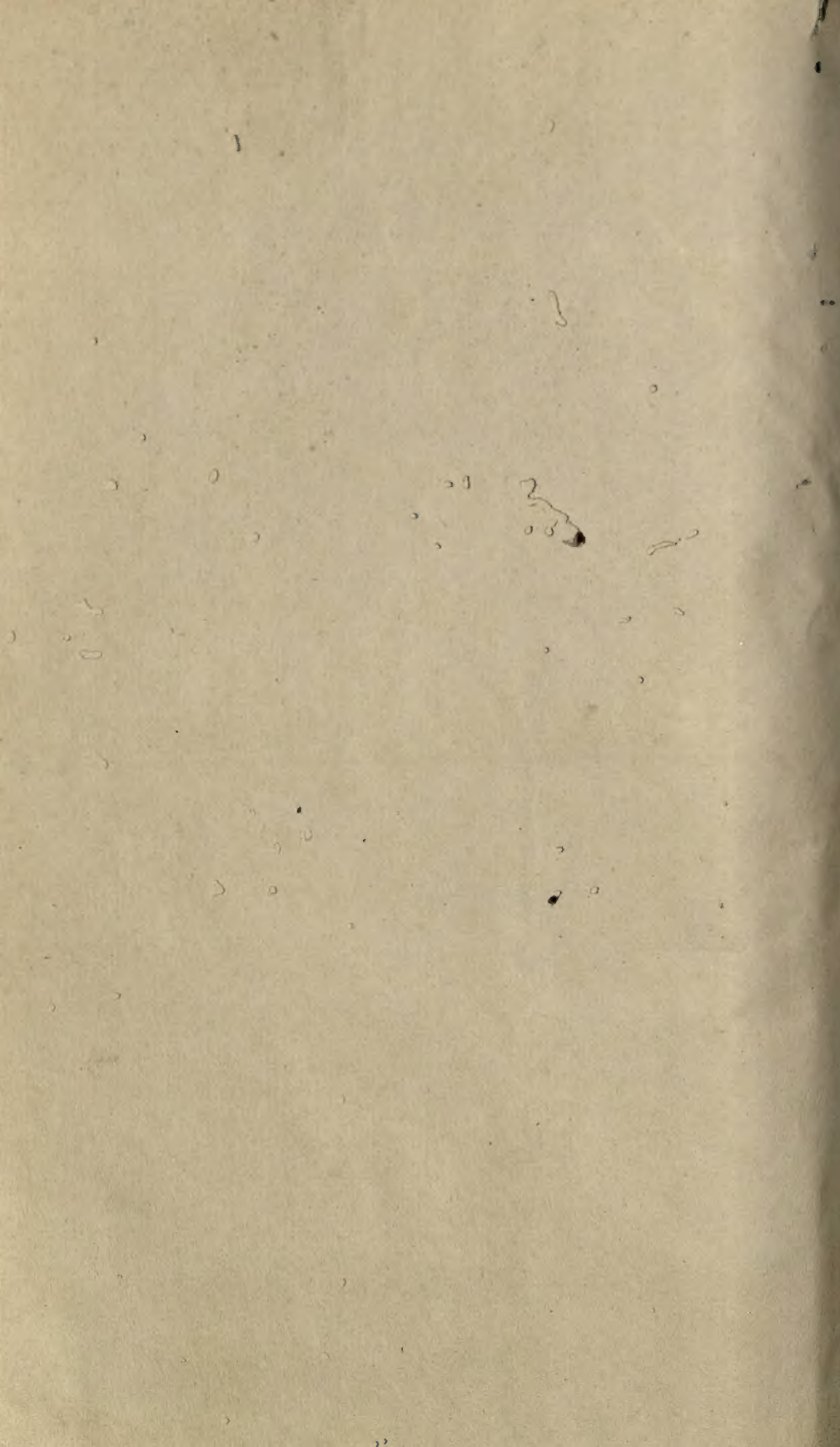




322
19/11/76



Journal of Personality and Social Psychology

Editor: DANIEL KATZ, *University of Michigan*
Associate Editor: ROBERT B. ZAJONC, *University of Michigan*

CONSULTING EDITORS

ROBERT P. ABELSON
Yale University
JOSEPH B. ADELSON
University of Michigan
GORDON W. ALLPORT
Harvard University
NORMAN H. ANDERSON
University of California, San Diego
ELLIOT ARONSON
University of Texas
ALBERT BANDURA
Stanford University
LEONARD BERKOWITZ
Cornell University
JAMES BIERI
University of Texas
JACK W. BREHM
Duke University
DONALD R. BROWN
University of Michigan
EUGENE BURNSTEIN
University of Michigan
ARNOLD H. BUSS
Rutgers University
LAUNO F. CARTER
Systems Development Corporation
ROSALIND D. CARTWRIGHT
*University of Illinois
Medical Center, Chicago*
DWIGHT W. CHAPMAN, JR.
Vassar College
IRVIN L. CHILD
Yale University
NICKOLAS B. COTTRELL
Kent State University
WILLIAM DEMBER
University of Cincinnati
MORTON DEUTSCH
Columbia University
SEYMOUR FESHBACH
University of Colorado

DONALD W. FISKE
University of Chicago
BASIL S. GEORGOPOULOS
University of Michigan
EDWIN P. HOLLANDER
*Tavistock Institute of Human Relations
London, England*
MURRAY HORWITZ
New York University
HERBERT C. KELMAN
University of Michigan
DANIEL J. LEVINSON
*Connecticut Mental Health Center
New Haven*
RICHARD A. LITTMAN
University of Oregon
ROBERT B. MACLEOD
Cornell University
MELVIN MANTIS
*Veterans Administration
Hospital, Ann Arbor, Michigan*
DAVID C. MCCLELLAND
Harvard University
CHARLES G. MCCLINTOCK
University of California, Santa Barbara
WILLIAM J. MCGUIRE
*Center for Advanced Study in the
Behavioral Sciences, Stanford, California*
VINCENT NOWLIS
University of Rochester
HAROLD L. RAUSH
University of Michigan
MILTON E. ROSENBAUM
University of Iowa
WILLIAM A. SCOTT
University of Colorado
JANET TAYLOR SPENCE
University of Texas
PERCY H. TANNENBAUM
University of Wisconsin
HARRY C. TRIANDIS
University of Illinois
JOSEPH VEROFF
University of Michigan

HELEN ORR VIRGINIA GONCI HENRY S. SALMON ROBERT D. GORMAN
Managing Editor Technical Editor Advertising and Business Manager Subscription Manager

VOLUME 5 1967

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Copyright © 1967 by the American Psychological Association, Inc.

CONTENTS OF VOLUME 5

ARTICLES, BRIEF ARTICLES

ALLPORT, GORDON W., & ROSS, J. MICHAEL. Personal Religious Orientation and Prejudice . . .	432
BANDURA, ALBERT, GRUSEC, JOAN E., & MENLOVE, FRANCES L. Some Social Determinants of Self-Monitoring Reinforcement Systems . . .	449
BANDURA, ALBERT, GRUSEC, JOAN E., & MENLOVE, FRANCES L. Vicarious Extinction of Avoidance Behavior . . .	16
BARUCH, RHODA. The Achievement Motive in Women: Implications for Career Development . . .	260
BERKOWITZ, LEONARD, & FRIEDMAN, PHILIP. Some Social Class Differences in Helping Behavior . . .	217
BERKOWITZ, LEONARD, & GEEN, RUSSELL G. Stimulus Qualities of the Target of Aggression: A Further Study . . .	364
BIERI, JAMES. See IRWIN, MARC.	
BIXENSTINE, V. EDWIN, & DOUGLAS, JOAN. Effect of Psychopathology on Group Consensus and Cooperative Choice in a Six-Person Game . . .	32
BLUM, GERALD S., GETWITZ, P. JAMES, & STEWART, CHARLES G. Cognitive Arousal: The Evolution of a Model . . .	138
BRYAN, JAMES H. See MIDLARSKY, ELIZABETH.	
BURKE, RONALD J. See MAIER, NORMAN R. F.	
BYRNE, DONN, GRIFFITT, WILLIAM, & STEFANIAK, DANIEL. Attraction and Similarity of Personality Characteristics . . .	82
CAIRNS, ROBERT B. Informational Properties of Verbal and Nonverbal Events . . .	353
COADY, HENRY. See SWINGLE, PAUL G.	
CONSTANTINOPLE, ANNE. Perceived Instrumentality of the College as a Measure of Attitudes toward College . . .	196
COWAN, GLORIA, & GOLDBERG, FAYE J. Need Achievement as a Function of the Race and Sex of Figures of Selected TAT Cards . . .	245
CRANE, WILLIAM D., & SCHRODER, H. M. Complexity of Attitude Structure and Processes of Conflict Reduction . . .	110
DEUTSCH, MORTON. See HORNSTEIN, HARVEY A.	
DILLEMAY, RONALD C. See LERNER, MELVIN J.	
DOUGLAS, JOAN. See BIXENSTINE, V. EDWIN.	
DUTTA, SATRAJIT, & KANUNGO, RABINDRA N. Retention of Affective Material: A Further Verification of the Intensity Hypothesis . . .	476
FEATHER, N. T., & SAVILLE, MARGARET R. Effects of Amount of Prior Success and Failure on Expectations of Success and Subsequent Task Performance . . .	226
FLANDERS, JAMES P., & THISTLETHWAITE, DONALD L. Effects of Familiarization and Group Discussion upon Risk Taking . . .	91
FOLEY, JEANNE M. See VERNON, DAVID T. A.	
FRIEDMAN, PHILIP. See BERKOWITZ, LEONARD.	
FRY, CHARLES L. A Developmental Examination of Performance in a Tacit Coordination Game Situation . . .	277
GALL, MEREDITH, & MENDELSON, GERALD A. Effects of Facilitating Techniques and Subject-Experimenter Interaction on Creative Problem Solving . . .	211
GARTNER, DOROTHY, & IVERSON, MARVIN A. Some Effects of Upward Mobile Status in Established and Ad Hoc Groups . . .	390
GEEN, RUSSELL G. See BERKOWITZ, LEONARD.	
GETWITZ, P. JAMES. See BLUM, GERALD S.	
GOLDBERG, FAYE J. See COWAN, GLORIA.	
GOLLOB, HARRY F., & LEVINE, JACOB. Distraction as a Factor in the Enjoyment of Aggressive Humor . . .	368
GORDON, ANDREW. See NISBETT, RICHARD E.	
GRIFFITT, WILLIAM. See BYRNE, DONN.	
GRUSEC, JOAN. See MISCHEL, WALTER.	
GRUSEC, JOAN E. See BANDURA, ALBERT.	
HAHN, SHWU CHING. See MARSHALL, HELEN R.	
HANCOCK, FRANCENA T. See SAMPSON, EDWARD E.	
HARMATZ, MORTON G. Verbal Conditioning and Change on Personality Measures . . .	175
HARNETT, DONALD L. A Level of Aspiration Model for Group Decision Making . . .	58
HARVEY, O. J. See WARE, ROBERT.	
HOFFMAN, MARTIN L., & SALTZSTEIN, HERBERT D. Parent Discipline and the Child's Moral Development . . .	45
HOLMES, DAVID S. Pupillary Response, Conditioning, and Personality . . .	98
HORNSTEIN, HARVEY A., & DEUTSCH, MORTON. Tendencies to Compete and to Attack as a Function of Inspection, Incentive, and Available Alternatives . . .	311
IRWIN, MARC, TRIPODI, TONY, & BIERI, JAMES. Affective Stimulus Value and Cognitive Complexity . . .	444
IVERSON, MARVIN A. See GARTNER, DOROTHY.	
JACKSON, DOUGLAS N. See STRICKER, LAWRENCE J.	
JELLISON, JERALD M., & MILLS, JUDSON. Effect of Similarity and Fortune of the Other on Attraction . . .	459
KAFER, JOHN, JR. See PYRON, BERNARD.	
KANUNGO, RABINDRA N. See DUTTA, SATRAJIT.	
KARLINS, MARVIN, & LAMM, HELMUT. Information Search as a Function of Conceptual Structure in a Complex Problem-Solving Task . . .	456

KATZ, DANIEL. Editorial	1
KATZ, PHYLLIS, & ZIGLER, EDWARD. Self-Image Disparity: A Developmental Approach	186
KLECK, ROBERT E., & WHEATON, JERRY. Dogmatism and Responses to Opinion-Consistent and Opinion-Inconsistent Information	249
KLEIN, MARJORIE H. Compliance, Consistent Conformity, and Personality	239
LAMM, HELMUT. See KARLINS, MARVIN.	
LAUGHLIN, PATRICE R. Incidental Concept Formation as a Function of Creativity and Intelligence	115
LEFCOURT, HERBERT M. Effects of Cue Explication upon Persons Maintaining External Control Expectancies	373
LEIDERMAN, P. HERBERT. See SHAPIRO, DAVID.	
LENER, MELVIN J., DILLEHAY, RONALD C., & SHERER, WILLIAM C. Similarity and Attraction in Social Contexts	481
LENER, MELVIN J., & MATTHEWS, GAIL. Reactions to Suffering of Others under Conditions of Indirect Responsibility	319
LESTER, DAVID. Suicide, Homicide, and the Effects of Socialization	466
LEVINE, JACOB. See GOLLOB, HARRY F.	
LIPETZ, MILTON E., & OSSORIO, PETER G. Authoritarianism, Aggression, and Status	468
LOEHLIN, JOHN C. Word Meanings and Self-Descriptions: A Replication and Extension	107
LOEW, CLEMENS A. Acquisition of a Hostile Attitude and Its Relationship to Aggressive Behavior	335
LOTT, ALBERT J. See PORIER, GARY W.	
MAIER, NORMAN R. F., & BURKE, RONALD J. Response Availability as a Factor in the Problem-Solving Performance of Males and Females	304
MANIS, MELVIN. Context Effects in Communication	326
MARSHALL, HELEN R., & HAHN, SHWU CHING. Experimental Modification of Dramatic Play ..	119
MATTHEWS, GAIL. See LENER, MELVIN J.	
MCCCLINTOCK, CHARLES G., & MCNEEL, STEVEN P. Prior Dyadic Experience and Monetary Reward as Determinants of Cooperative and Competitive Game Behavior	282
MCNEEL, STEVEN P. See MCCCLINTOCK, CHARLES G.	
MENDELSON, GERALD A. See GALL, MEREDITH.	
MENLOVE, FRANCES L. See BANDURA, ALBERT.	
MESSICK, SAMUEL. See STRICKER, LAWRENCE J.	
MEYER, GÜNTHER. See RIEGEL, KLAUS F.	
MIDLARSKY, ELIZABETH, & BRYAN, JAMES H. Training Charity in Children	408
MILLS, JUDSON. See JELLISON, JERALD M.	
MISCHEL, WALTER, & GRUSEC, JOAN. Waiting for Rewards and Punishments: Effects of Time and Probability on Choice	24
NISBETT, RICHARD E., & GORDON, ANDREW. Self-Esteem and Susceptibility to Social Influence	268
OSSORIO, PETER G. See LIPETZ, MILTON E.	
PAPAGEORGIS, DEMETRIOS. Anticipation of Exposure to Persuasive Messages and Belief Change	490
PORIER, GARY W., & LOTT, ALBERT J. Galvanic Skin Responses and Prejudice	253
PYRON, BERNARD, & KAFER, JOHN, JR. Recall of Nonsense and Attitudinal Rigidity	463
PYTKOWICZ, ANN ROTH, WAGNER, NATHANIEL N., & SARASON, IRWIN G. An Experimental Study of the Reduction of Hostility through Fantasy	295
RHINE, RAMON J. The 1964 Presidential Election and Curves of Information Seeking and Avoiding	416
RICHARDS, BARRIE. See SMITH, KAY H.	
RIEDEL, KLAUS F., RIEGEL, RUTH M., & MEYER, GÜNTHER. A Study of the Dropout Rates in Longitudinal Research on Aging and the Prediction of Death	342
RIEDEL, RUTH M. See RIEGEL, KLAUS F.	
RODRIGUES, AROLD. Effects of Balance, Positivity, and Agreement in Triadic Social Relations	472
ROSENHAN, DAVID, & WHITE, GLENN M. Observation and Rehearsal as Determinants of Prosocial Behavior	424
ROSS, J. MICHAEL. See ALLPORT, GORDON W.	
RUBIN, IRWIN M. Increased Self-Acceptance: A Means of Reducing Prejudice	233
SALTZSTEIN, HERBERT D. See HOFFMAN, MARTIN L.	
SAMPSON, EDWARD E., & HANCOCK, FRANCENA T. An Examination of the Relationship between Ordinal Position, Personality, and Conformity: An Extension, Replication, and Partial Verification	398
SARASON, IRWIN G. See PYTKOWICZ, ANN ROTH.	
SAVILLE, MARGARET R. See FEATHER, N. T.	
SCHRODER, H. M. See CRANO, WILLIAM D.	
SCHULMAN, JEROME L. See VERNON, DAVID T. A.	
SGAN, MABEL L. Social Reinforcement, Socioeconomic Status, and Susceptibility to Experimenter Influence	202
SHAPIRO, DAVID, & LEIDERMAN, P. HERBERT. Arousal Correlates of Task Role and Group Setting	103
SHERER, WILLIAM C. See LENER, MELVIN J.	
SMITH, CAROLE R., WILLIAMS, LEV, & WILLIS, RICHARD H. Race, Sex, and Belief as Determinants of Friendship Acceptance	127
SMITH, KAY H., & RICHARDS, BARRIE. Effects of a Rational Appeal and of Anxiety on Conformity Behavior	122
STAGER, PAUL. Conceptual Level as a Composition Variable in Small-Group Decision Making	152
STEFANIAK, DANIEL. See BYRNE, DONN.	
STEWART, CHARLES G. See BLUM, GERALD S.	
STONE, WILLIAM F. Autokinetic Norms: An Experimental Analysis	76
STRICKER, LAWRENCE J., MESSICK, SAMUEL, & JACKSON, DOUGLAS N. Suspicion of Deception: Implications for Conformity Research	379

SWINGLE, PAUL G., & COADY, HENRY. Effects of the Partner's Abrupt Strategy Change upon Subject's Responding in the Prisoner's Dilemma	357
THISTLETHWAITE, DONALD L. See FLANDERS, JAMES P.	
TRIPODI, TONY. See IRWIN, MARC.	
VERNON, DAVID T. A., FOLEY, JEANNE M., & SCHULMAN, JEROME L. Effect of Mother-Child Separation and Birth Order on Young Children's Responses to Two Potentially Stressful Experiences	162
WAGNER, NATHANIEL N. See PYTKOWICZ, ANN ROTH.	
WARE, ROBERT, & HARVEY, O. J. A Cognitive Determinant of Impression Formation	38
WATTS, WILLIAM A. Relative Persistence of Opinion Change Induced by Active Compared to Passive Participation	4
WEINSTOCK, ALLAN R. Family Environment and the Development of Defense and Coping Mechanisms	67
WHEATON, JERRY. See KLECK, ROBERT E.	
WHITE, GLENN M. See ROSENHAN, DAVID.	
WILLIAMS, LEV. See SMITH, CAROLE R.	
WILLIS, RICHARD H. See SMITH, CAROLE R.	
WOLITZKY, DAVID L. Cognitive Control and Cognitive Dissonance	486
WRIGHT, LOGAN. Factor Structure of Duncan's Personality Integration Scale	348
WUEBBEN, PAUL LANE. Honesty of Subjects and Birth Order	350
ZIGLER, EDWARD. See KATZ, PHYLLIS.	

EDITORIAL

It is a pleasure to report to our readers and authors that the mounting publication lag, which reached a peak during early 1966 of 18 months, dropped to 12 months by the end of that year. Moreover, it should remain at that level for 1967. Publication lag, it should be remembered, is not identical with editorial lag. The editorial process can be handled within 4 months for our Journal. The additional time before publication is due to the acceptance of more manuscripts than pages allocated in that accepted manuscripts pile up awaiting their turn as space becomes available. Even though the same high rejection rate is maintained, publication lag will increase if the yearly allocation of pages does not keep pace with the increased inflow of manuscripts. The publication delay in late 1965 and early 1966 was primarily an organizational lag in awarding adequate pages for the anticipated growth of our Journal. This was remedied in 1966 as the decisions of the Publications Board became implemented by the Board of Directors.

Another factor contributed to the 1966 reduction in publication delay. The rate of increase of submitted manuscripts declined for two reasons. One was the lag itself. As the time between submission of manuscripts and their publication goes up, authors turn to other publication outlets. The second was the appearance of new journals under private auspices to which authors could send their papers. The first reason was by far the more important. As the months went by and the publication lag was brought under control, the word quickly got around, and during the latter part of the year the input began to rise again. This means that the problem of sufficient pages will remain with us in the foreseeable future.

The prospects are excellent, however, that the major aspects of the problem can be satisfactorily handled in that we have achieved an organizational solution in methods for dealing with the issues. The Board

of Directors and the Council of Representatives have approved the granting of considerable autonomy to the Publications Board for major decisions about the journals including the allocation of pages. So long as the Publications Board stays within the income derived from the American Psychological Association journal operation, it can make provision for increasing the size of the hard-pressed journals. Recent improved methods of accounting in the APA central office show that the journals are and have been self-supporting and generally contribute small amounts to the "black side of the ledger." Moreover, the new policy will provide for more volumes per year for a given journal as the space needs increase. With each additional volume, the price to subscribers goes up, and the market as well as the inflow of manuscripts will help to determine the growth of the journals. This policy will not be applied to penalize the needs of an important journal with a relatively small circulation since the journal operation as a whole, rather than a specific publication, must break even. What it means is that a journal with a wide audience in a growing field can publish adequate volumes and even contribute greater amounts to the support of more specialized periodicals.

The move toward giving the Publications Board more autonomy is a significant step in the decentralization of decision making within the APA structure. It augurs well for the future of the APA journals. The facts are that the Board of Directors with its overload of complex, difficult problems does not have the time for thorough consideration of publication matters without sacrificing attention from other pressing issues. The Publications Board has representatives from the Council of Editors and can also avail itself of the recommendations and reports from this Council. Hence the Publications Board is the most knowledgeable group about the journals and is in a better position to make operational decisions within the broad framework of the

APA policy than any other body in the APA. Moreover, with this as its sole task, it can stay on top of a problem and make decisions more promptly than would be possible through referral to the Board of Directors.

In anticipating the needs for publication in the two most crowded journals, the Publications Board has recommended, and the Council of Representatives has approved, an additional volume of the *Journal of Personality and Social Psychology* in 1967 to make it a three volume a year periodical and an additional volume for the *Journal of Experimental Psychology* to make it a three volume a year journal in 1967 and perhaps a four-volume journal in 1968.

Another decision also approved by the Council of Representatives in its August 1966 meeting was to change the character of the publication of the *Psychological Monographs* by distributing its pages to seven other journals. Thus the *Journal of Personality and Social Psychology* will issue a monograph supplement for longer articles to the extent of its share of monograph space. It will also have to match these added pages by pages of its regular quota. Specifically, 70 pages have been allocated to our Journal which will permit the publication of a few monographs during 1967. These will go to all subscribers of our journal without extra charge during 1967. Subscribers to the *Monographs* will receive the supplements from all seven journals so that, for library purposes as well as for the needs of the generalist, the separates can be bound to constitute a volume which will be a continuation of the old *Psychological Monographs*.

The present editorial staff has now started its lame-duck year in that all new manuscripts should go to the editor-elect, William McGuire, 604 Mathematics Building, Columbia University. The old editor and his associates will be occupied with the review and processing of manuscripts received during 1966.

During the past year we have continued the policy of attempting a broad coverage of the field of social psychology and personality. Since the *Journal of Personality and Social Psychology* is an APA journal, we believe that we cannot narrow its focus to a limited

set of problems, but must accept good manuscripts representative of the diverse areas of research of workers in the field. Thus we have published papers which deal with the central domain of interpersonal behavior, group process and group outcomes, and we have also included studies which deal with essentially individual processes but are social in content. For example, dissonance studies are primarily concerned with the effects of cognitive conflict no matter how produced, but the variables manipulated and the effects studied are often social in substantive meaning. In other words, we have been concerned both with the uniquely social and the application of individual psychology to social problems. The justification is that social psychologists and personality theorists are involved in research directed at social variables whether the interest is in social factors as an independent or as a dependent variable. Hence the studies considered appropriate for our Journal have ranged a wide gamut from investigations of intergenerational cultural change to physiological correlates of social stress. We have modified this policy in one essential respect, namely, in the case of papers for which there is already a specialized journal especially relevant to the article in question. Where the study deals, for example, with an understanding of psychopathology, we refer it to the *Journal of Abnormal Psychology* even though its content is appropriate for social psychology or personality theory. Similarly, in the field of verbal learning, we believe the specialized journals in this area have a prior claim to manuscripts concerned with verbal conditioning.

A journal of the size and complexity of the *Journal of Personality and Social Psychology* can operate effectively only with the aid of a highly competent staff of reviewers representing the many specialized interests in the field. We are indebted to our editorial board for their continued dedication to the task of evaluating manuscripts. We also wish there were some way of giving adequate recognition to the following consultants who have given generously of their time and thought in reviewing two or more papers during the past year:

REVIEWERS IN 1966

Jack Atkinson	David C. Glass	Gerald Lesser	Hayne W. Reese
Kurt W. Back	Lewis G. Gillberg	George Levanger	Morton Rosenberg
Edwin Bixenstine	Michael Goldstein	David Lewis	Seymour Rosenberg
Dana Bramel	Joel Greenspoon	David Marlowe	Julius B. Rotter
Timothy C. Brock	John Hagen	David McNeill	Edward F. Sampson
Donn Byrne	John Harding	Martha Mednick	Irwin G. Sarason
J. Merrill Carlsmith	Martin Hoffman	Leo Meltzer	Kenneth Schneider
Dorwin Cartwright	Jack Hukanson	Gerald A. Mendelsohn	Paul F. Secord
Albert M. Chaimmah	Philip S. Holzman	Charles Metzner	Anthony J. Smith
Natalia P. Chapanis	Raymond Hunt	John Milledand	Charles P. Smith
Emory L. Cowen	Douglas N. Jackson	Norman Miller	Diedrick Snook
Douglas P. Crowne	Stanislav Kasi	Judson R. Mills, Jr.	Arthur W. Staats
Robyn Dawes	Solis L. Kates	Walter Mischel	Ivan D. Steiner
Richard deCharms	Irwin Katz	Robert W. Moulton	Leonard Ullmann
Clinton B. De Soto	Charles A. Kiesler	Paul H. Mussen	William S. Verplanck
Donald Dorfman	Nathan Kogan	Thomas Natanson	William A. Watts
Norman S. Endler	Samuel S. Komorita	Stuart Oskamp	Karl E. Weick
Jonathan Freedman	Leonard Krasner	Marc Pilisuk	Daniel J. Weintraub
Elizabeth G. French	Peter J. Lang	Alfred C. Raphelson	Edward Zigler
Philip S. Gallo	Richard Lazarus	Bertram Raven	Philip G. Zimbardo

DANIEL KATZ

RELATIVE PERSISTENCE OF OPINION CHANGE INDUCED BY ACTIVE COMPARED TO PASSIVE PARTICIPATION¹

WILLIAM A. WATTS

University of California, Berkeley

This study examined the relative persistence of opinion change induced by active participation (writing an argument) versus passive participation (reading an argument). Through prestudy and modification of the persuasive messages, the immediate effects of the active and passive conditions were equated, but, over time (6 weeks), the active participation condition showed clear temporal superiority ($F = 6.84$, $p < .05$). Active participation also resulted in greater involvement (subsequent discussion of, and reading about, the topic) and superior recall of the topic and side supported. Involvement was positively related to persistence of opinion change. Recall was complexly related to persistence, being positive in the writing conditions, but negative in the reading conditions, and the order of measurement of opinion and recall (opinion recall vs. recall opinion) on the delayed questionnaire had a significant effect with the recall-first order facilitating agreement.

The present study was designed to investigate the relative persistence of opinion change induced by improvising and writing an argument advocating an attitudinal position (active participation) compared to reading a persuasive message (passive participation) which argues in favor of the same attitudinal position. Both the method of opinion induction and the temporal persistence of induced opinion change were further related to subjects' memory of the topic, and side taken in the induction session and to three indexes of their involvement with the topic. Hence, three issues are of concern: the overall persistence of opinion change conferred by each method of induction, the effects of each induction method upon the subjects' memory and involvement, and, finally, the relationships that both memory and involvement bear to persistence of opinion change.

The immediate effects of active participation in the form of role playing or writing persuasive essays have been much studied, both under conditions where this participation accompanied the assimilation of new information through reading or hearing a persuasive message (e.g., Janis & King, 1954; Kelman, 1953; King & Janis, 1956) and where no new information was presented, the subject being

called upon to improvise arguments (Cohen, Brehm, & Fleming, 1958; Cohen & Latané, 1962; Elms & Janis, 1965; Janis & Gilmore, 1965; Rabbie, Brehm, & Cohen, 1959; Rosenberg, 1965). With few exceptions (e.g., Stanley & Klausmeier, 1957) this technique has been shown to be effective in inducing attitude change, although there is some disagreement as to the nature of this self-persuasion process (e.g., Brehm, 1965; Elms & Janis, 1965).

The question under examination in the present research, the temporal persistence of such induced change, has been relatively neglected. The few studies that do bear upon this question were not primarily designed to investigate temporal effects, and little attempt was made to study the variables that might mediate opinion retention.

In their classic studies of group discussion and decision Lewin (1958) and his co-workers found that the amount of compliance induced by this technique was greater after 4 weeks than after 2, although some increments also occurred in the "individual instruction" conditions. Lewin conjectured that the greater effectiveness of group discussion at both time intervals was due in part to differences in degree of involvement.

Culbertson (1957) found significant change after a delay of 7-10 days both for subjects who had role played and for observers com-

¹ This study was supported by University of California research funds. Thanks are extended to David Whittaker for his assistance in collecting the data.

pared to controls. Furthermore, those subjects who had participated in role playing showed greater change than the observers and scored significantly higher on six indexes of involvement than did the observers. However, since no immediate measures of opinion change were obtained, it is impossible to assess the temporal effects. Similarly, Janis and Mann (1965) found emotional role playing (enacting the role of a patient just diagnosed as having lung cancer) to be more effective than listening to a tape recording of a similar role-playing session in changing subjects' attitudes toward smoking. This superiority was evidenced not only in immediate attitude change, but also in the subjects' reported decrease in cigarette consumption when followed up 2 weeks later.

Mitnick and McGinnies (1958)* reported greater retention of induced opinion change among groups who had seen a film and then spent 30 minutes in informal discussion of it and related issues than among groups who had only seen the film. Immediately after the treatments there was no difference in attitude change between the two conditions, but when measured again a month later the film-alone groups had regressed significantly back toward their pretreatment position, whereas the film-discussion groups showed no significant change from their immediate posttreatment scores. A 33-item multiple-choice test of factual information presented in the film indicated that there were no significant differences between active participants in the discussion and those subjects who had seen the film alone in either initial learning or retention, although they did learn significantly more than passive participants who had not entered into the discussion.

These studies which bear upon persistence of opinion change induced by group discussion and role playing all involved, in varying degrees, the presentation of new information in the process. The experimental paradigm employed in the present study differed in the important respect that subjects in the active-writing condition had no opportunity to gain new information during the opinion-induction session. Thus, it might be argued that change resulting from improvisation only would be quite transitory as compared to a reading in-

duction where opinion change presumably results from the assimilation of new information into the cognitive system.

On the other hand, there is evidence that subjects who have actively participated are more involved with the topic (Culbertson, 1957) and more confident of their opinions (Janis & King, 1954). One would expect that with greater degrees of improvisation, reaching an apex in studies like the present one where the subject is given no new information but called upon to invent arguments, there would be correspondingly greater degrees of personal involvement with the task.

Involvement, which has been an elusive and ill-defined concept in psychology, was operationally defined in the present study in terms of the subjects' subsequent discussion of the topic, reading about the topic in other sources, and their degree of interest in the task. These particular indexes of involvement were selected because on an a priori basis they would appear likely to be related to persistence of opinion change. In the event of such activity, there is little reason to believe that subjects in the active-participation condition would rapidly revert to their initial opinion level, unless the new position was clearly counter-norm where much of the information they might receive would oppose their newly held opinion. Consequently, given approximately equal initial change, greater persistence of opinion change and more subsequent discussion and reading about the topic would be predicted for those subjects who composed a persuasive argument.

Furthermore, at a delayed interval one might expect subjects in the active-participation condition to be better able to recall the issue and the side supported than their counterparts who had read a persuasive passage. Research on the relationship between active participation and learning and retention of content (e.g., Kurtz & Hovland, 1953; Mitnick & McGinnies, 1958) is largely irrelevant here since the subjects received no new information except the assigned title and side that they were to support. Rather, the prediction of superior recall of these aspects of the situation is predicated upon the notion of greater involvement and subsequent seeking of additional information through discussion

and reading in the active-participation condition.

In addition, one might speculate that the order of measurement of opinion and recall (i.e., opinion recall versus recall opinion) in the delayed questionnaire would have an effect so that being asked to recall the topic and side taken prior to stating opinion would facilitate agreement by reminding the subjects of the arguments they had previously read or written. Such an effect might be quite similar to that of restating a positive source when measuring subjects' opinions at a delayed interval (e.g., Kelman & Hovland, 1953) since the purported purpose of the opinion-induction session was quite respectable (see the Method section). This effect should, however, be greater for subjects who had written a persuasive argument than for those who had read a message because of stronger commitment to the attitudinal position acquired by actively defending it.

METHOD

Subjects

The subjects were 140 students enrolled in an introductory course in educational psychology at the University of California, the majority of whom were females. All subjects participated in the study during the regular class meetings. Six sections of the course were used, ranging in size from 16 to 40. The precise number of subjects serving in each experimental condition is shown in each cell of tables 1 through 5.

Since a repeated-measurements design was employed in which each subject was measured twice (to be discussed), there were inevitably a number of subjects who were absent on one of these occasions and, consequently, could not be included in the data analyses. Of the 174 subjects attending the first session 141 were also present for the second. One subject who failed to complete the entire questionnaire was excluded leaving the final *N* of 140.

Procedure

The study was represented to the subjects as an attempt to develop instruments to measure analytical thinking ability in future teachers. Those subjects who were to read a persuasive message were told that they would be asked to read a standardized passage dealing with a controversial issue and that their task was to select and underline in each paragraph the shortest clause which epitomized the whole point being made. They were then allowed 8 minutes for reading and underlining in the prescribed manner one of the persuasive messages.

Subjects assigned to the condition of writing an argument were instructed as follows:

On the following page you will find a statement concerning a controversial issue. Your task is to organize your thoughts and write a strong convincing argument supporting the side of the issue indicated; thus, for example, if the statement were, "Capital punishment should be abolished," your task no matter how you feel about it personally would be to write the most convincing exposition possible pointing out the reasons you can think of for abolishing capital punishment.

Subjects in this condition were also allowed 8 minutes, which proved sufficient, to write their arguments.

Within each of the opinion-induction conditions three different issues were used with each group of subjects reading or writing, as the case might be, about only one issue. These six conditions were randomly assigned to the aforementioned six sections of the educational psychology course. Immediately after the opinion-induction session, subjects filled out an opinion questionnaire on all three issues. Hence, each person served as an experimental subject on one issue and as a control on the other two.² That is, the magnitude of opinion change for subjects who had read or written about a particular issue was ascertained by comparing their mean opinion score to that of groups who had read or written about the other two issues and whose opinion on the issue in question had not been subjected to any persuasive influence.

The subjects were thanked for participating, and no further mention was made of the study for 6 weeks. At that time the experimenter returned and asked the students to fill out the opinion questionnaire a second time. The instructions were as follows:

Some time ago you rated the extent of your agreement with a number of statements concerning current issues. Today we would like you to check them again so as to see whether the population's opinion on these issues is changing over time, or, instead, is relatively stable, as well as to provide some additional information.

On this questionnaire the subject was not only asked to state his opinion, but also to recall and state the topic and side taken on the issue in the previous induction session. The order of presentation of the opinion and recall items was varied in the questionnaires with half of the subjects being first asked to state their opinion, having no knowledge that they would subsequently be asked to recall the material, while the other half were first asked to

² Due to an administrative error, a number of subjects who should have provided controls for the Secretary of State issue received a form of the questionnaire which did not contain these items. The items concerning the other two issues were identical in both forms, and there were no differences in means between groups receiving the two-issue or three-issue form of the questionnaire. It did, however, considerably reduce the size of the control group for the Secretary of State issue (see Table 1).

recall the topic and side taken and then to state their opinion. Subjects were instructed not to look ahead in the booklet, but to fill out each page completely before proceeding to the next page and not to look back and review their work. Within each of the six sections subjects were randomly assigned to one of the order of measurement conditions by alternating the forms of the questionnaire booklets. The last page of each booklet contained three questions designed to measure the subjects' degree of involvement with the issue.

Thus, two opinion scores were obtained on each subject, one immediately after the initial induction session and another 6 weeks later, while only one score was obtained on the recall and involvement questions, and this was on the delayed measure.

Materials

The three issues required by the present design were taken from a set of four that had been used in a previous study (Watts & McGuire, 1964). These issues were: "Puerto Rico should be admitted to the Union as the 51st state"; "Courts should deal more leniently with juvenile delinquents"; and "The Secretary of State should be elected by the people, not appointed by the President." Since it was most desirable from a statistical standpoint, in order to avoid troublesome regression artifacts, that the immediate mean opinion change induced by the persuasive message on each issue be approximately equal to that induced by writing an argument on the same issue, prestudy was carried out on comparable educational psychology students, and the persuasive messages were modified so as to meet this criterion as closely as possible. (As will be seen in the Results section, this effort was quite successful.³)

Each persuasive message argued in favor of one of the above statements and consisted of a title and about 600 words of text divided into four paragraphs. The introductory paragraph stated that the communication favored the statement and mentioned three arguments supporting it. Each of the following paragraphs developed in detail one of these three arguments. The messages were written in a calm, authoritative fashion with frequent citation of (purported) facts and figures bearing on the point being made.

The materials given to those students assigned to write a convincing argument consisted of a statement which was identical to the title of one of the articles received by the other groups (e.g., "Courts should deal more leniently with juvenile delinquents") and

two sheets of blank paper on which the subject was to write his argument.

The first opinion questionnaire consisted of a series of nine statements, three pertaining to each of the three issues involved. Opinion was measured by having the subject check his degree of agreement with each statement on a 15-point graphic scale which followed the statement. The left end of each scale was labeled "Definitely Disagree" and the right end labeled "Definitely Agree." Two of the three statements pertaining to each issue were worded so that "Definitely Agree" responses indicated extreme agreement with the position advocated in the issue, while the third statement was worded in reverse with extreme disagreement with the statement corresponding to the position advocated. For each issue the three items were summed, with appropriate reflection of items keyed in reverse, so as to provide an opinion score for each subject on each of the three issues. Hence, each score had a possible range of from 3 to 45. These items were not presented in triads, but were intermixed throughout the questionnaire among those pertaining to the other issues utilized.

The final questionnaire, administered 6 weeks after the initial session, contained the same opinion items and, in addition, asked the subject as a recall measure to state in his own words the general topic and the side taken in the previous session. Alternate forms of the questionnaire were prepared with these recall questions preceding the opinion items in half of the booklets and following them in the other half. The last page of each questionnaire contained three questions designed to measure the degree of involvement the subject felt with the issue, providing 15-point rating scales for their responses. These were "How interesting did you find the task of reading or writing about the issue assigned?" (1, "not at all interesting," 15, "very interesting"); "Have you discussed the topic with anyone since the time that you wrote the essay or read the article?" (1, "not at all," 15, "very often"); and "Have you read anything else about the topic since the first time you participated in the study?" (1, "nothing at all," 15, "very often"). This last question was followed with "If yes, where? (e.g., newspaper, magazine, etc.)"

RESULTS

Before discussing the temporal effects, it should be established that opinion change initially occurred in each of the induction methods. This was done by separately comparing the means for both the reading and writing inductions to their control for each issue. It can be seen in Table 1 that, by and large, both methods induced a significant amount of opinion change on the immediate-after measures. Only one of the six comparisons (the writing condition on the Puerto

³ Actually, the success of this attempt to equate initial opinion change exceeded expectations. Since greater persistence was predicted for the active-participation group, the prestudy efforts were primarily concerned with ensuring as much or greater change in the active as in the passive conditions to rule out the alternate interpretation of such temporal effects in terms of regression. Fortunately, within each of the three issues the means of the two induction methods were almost identical.

Rico issue) falls short of significance at the .05 level using Dunnett's (1955) method for multiple comparisons with a control. Since the temporal analyses were performed across the three issues in each induction method, it can be safely concluded that opinion change was initially induced in both the active and passive conditions.

The six classes utilized in the study were all sections of the same course and, therefore, were probably comparable; although, in the absence of pretreatment measures, one cannot be certain. There are, however, two relevant analyses that dispel most doubts. Since each experimental group also served as a control for the other issues, it is possible to separate the control means presented in Table 1 into their component classes. For example, the control mean for the Puerto Rico issue of 22.38 (see Table 1) is based upon four classes: one reading and one writing about the Secretary of State issue and their two counterparts in the juvenile delinquency issue. If the classes do not differ on the control issues, it is reasonable to assume that they were comparable on the experimental issues. When the effect of classes within each control is tested by analysis of variance, the largest F value obtained is .90, with $df = 3/90$. Hence, from this evidence, there is no cause for worry about lack of comparability among classes. The other attestation to their comparability is the absence of interaction between issues, and hence classes, with the experimental conditions (to be discussed subsequently).

It is apparent (see Table 1) that within each issue the means for the reading and writing inductions are of comparable magnitude. Since subjects in each of the above conditions had been randomly assigned to one of the two order of measurement conditions (opinion recall versus recall opinion) on the final questionnaire, the immediate post-induction scores were further partitioned on this basis and analyzed by a $3 \times 2 \times 2$ analysis of variance.⁴ The only significant source of variance was the main effect of issues ($F = 15.86$, $p < .01$, $df = 2/128$) with all other F ratios for main effects and their interactions being less than 1. The main effect of method of induction yielded a trivial F of .11 indicating that the labor expended pretesting and modifying the issues so as to obtain maximum equality on the immediate measures was not in vain.

Effects of Method of Induction and Order of Measurement on Persistence of Induced Opinion Change

Since the immediate postinduction opinion levels of the two induction conditions and the two orders of measurement were established as comparable, the analysis of the temporal effects was performed on the difference scores (i.e., the difference between the immediate postinduction score and the 6-week-delay score for each subject). These scores were

⁴ All analyses of variance involving unequal cell frequencies were computed by the General Linear Hypothesis Program of the Health Sciences Computing Facility of UCLA.

TABLE 1
INITIAL EFFECTIVENESS OF EACH METHOD OF OPINION INDUCTION COMPARED TO CONTROLS

Issues	Controls	Reading	Writing	Significance of comparisons
Puerto Rico should be admitted to the Union as the 51st State.	22.38 ($N = 79$)	27.32 ($N = 40$)	26.52 ($N = 21$)	R vs. C, $t = 2.84^{**}$ W vs. C, $t = 1.82^{*}$
The Secretary of State should be elected by the people, not appointed by the President.	10.24 ($N = 25$) ^a	20.76 ($N = 17$)	21.12 ($N = 16$)	R vs. C, $t = 3.07^{**}$ W vs. C, $t = 3.25^{**}$
Courts should deal more leniently with juvenile delinquents.	28.82 ($N = 94$)	33.42 ($N = 19$)	34.07 ($N = 27$)	R vs. C, $t = 2.13^{**}$ W vs. C, $t = 2.88^{**}$

Note.—The significance levels of all comparisons reported in this table were estimated by Dunnett's (1955) method for comparing multiple treatment means with a control. The numbers in parentheses in this and subsequent tables indicate the number of scores on which the cell means are based.

^a The smaller N in this cell is explained in Footnote 2.

* $p < .10$.

** $p < .05$.

analyzed in a $3 \times 2 \times 2$ analysis of variance in which the three issues were considered as a variable. Inasmuch as no significant effects due to issues nor to any of the Issues \times Treatments interactions were found (all yielded F ratios < 1), the means presented in Table 2, and subsequent tables, have been collapsed across the three issues.

It is clear from the data in Table 2 that the method of induction has a strong effect upon the temporal persistence of opinion change ($F = 6.84$, $p < .05$). As predicted, subjects in the active-participation condition, who improvised persuasive arguments, show greater persistence of the initially induced opinion change over the 6-week interval than do their counterparts who read a persuasive message.

In fact, those subjects in the active-participation condition show a delayed persuasive effect or " sleeper effect " in that they became more favorable, although not significantly so, toward the position advocated over the 6-week time interval increasing a mean of 1.44 points from their postinduction level ($t = 1.50$, $p < .20$). Their counterparts who had read a persuasive message retrogress a mean of 1.59 points, toward the control level, over

the same time period. Hence, after 6 weeks the active-participation group differs significantly from the control ($t = 3.60$, $p < .01$), whereas the passive reading group, having lost 48% of the initially induced change, is no longer significantly different from the control level ($t = 1.13$). These data parallel the trends obtained by Mitnick and McGinnies (1958) for the effects of discussion of a film upon persistence of opinion change. The temporal superiority of active over passive participation was even greater in the present study in that the trend for the "active" group was a delayed increase in agreement despite the fact that no group interaction was involved in the induction session, nor was any new information presented.

It is not unreasonable to think that subjects who had written arguments would continue to rehearse and to think of additional arguments that they could have used for some time after the experimental session, particularly considering the brief period of time (8 minutes) that they were given to compose their arguments. Such rehearsing and a greater tendency on the part of active participants for subsequent discussion and reading about the topic (to be discussed) can readily explain the delayed-action effect obtained.

Order of measurement also had a significant main effect ($F = 3.91$, $p = .05$) with those subjects who were asked to recall the topic and side taken prior to stating their opinion showing greater agreement with the opinion statements. This effect is interpreted in terms of the recall-first order serving to remind the subjects of the arguments that they had previously read or written. As such, it is somewhat analogous to the facilitating effect of restating a positive source at the time of delayed opinion measurement (e.g., Kelman & Hovland, 1953). If this is correct, one would expect the advantage of the recall-first order to be limited to positive, respectable "sponsorships" (Janis & Gilmore, 1965) and message sources. Indeed, for the case of negative or suspect sponsorships and sources, the recall-first order should produce a decrement similar to that of restating a negative source (Kelman & Hovland, 1953).

It was also predicted that the facilitating

TABLE 2
PERSISTENCE OF OPINION CHANGE AS RELATED TO
METHOD OF INDUCTION AND ORDER
OF MEASUREMENT

Opinion-induction condition	Order of measurement on delayed questionnaire		<i>M</i>
	Opinion 1st	Recall questions 1st	
Writing			
Postinduction opinion mean	29.22	27.50	28.36
Subsequent change	-0.38 (<i>N</i> = 32)	+3.25 (<i>N</i> = 32)	+1.44 (<i>N</i> = 64)
Reading			
Postinduction opinion mean	27.67	27.10	27.38
Subsequent change	-2.05 (<i>N</i> = 37)	-1.15 (<i>N</i> = 39)	-1.59 (<i>N</i> = 76)
Overall			
Postinduction opinion mean	28.37	27.28	27.83
Subsequent change	-1.28 (<i>N</i> = 69)	+0.83 (<i>N</i> = 71)	-0.21 (<i>N</i> = 140)

effect of the recall-first order of measurement would be greater for subjects who had written arguments because it would remind them of their greater behavioral commitment to the attitudinal position. While the obtained trend is in the predicted direction, this interaction effect between method of opinion induction and order of measurement falls far short of statistical significance ($F = 1.54$).

Recall of the Topic and Side Taken Related to Method of Induction and Opinion Change

In discussing the order of measurement of recall and opinion in the previous section, no distinction was made between those subjects who were able and those who were unable to recall these aspects of the opinion-induction session when asked to do so. It would seem reasonable to expect that at a delayed interval those subjects who had improvised arguments would be better able to recall the topic and side supported (these being the only aspects of the issues constant across the reading and writing conditions) than their counterparts who had read a communication. This supposition was based upon the notion that subjects who had improvised arguments would be more personally involved with the issues and more likely to seek additional information subsequent to the induction session. Strong support was obtained for this conjecture with 88% of the subjects who had written arguments correctly stating the topic and side taken⁵ in the initial session 6 weeks earlier compared to 43% in the reading-induction condition. This difference in percentages is highly significant ($Z = 5.49$, $p < .01$).

The fact that the differential recall rates for the two induction conditions parallel the differential changes in opinion over time might suggest that they are functionally dependent upon one another, with the more rapid decay of opinion in the reading condition being due to the higher rate of forgetting. In order to determine the relationship between memory of these basic aspects of the issue and persistence of opinion change, subjects' opinion scores were further partitioned on the basis of

⁵ Since only 14% of the subjects recalled one of these aspects of the message (topic or side taken) without recalling the other, the two were combined for these analyses.

TABLE 3
PERSISTENCE OF OPINION CHANGE AS RELATED
TO ABILITY TO RECALL TOPIC AND SIDE
TAKEN AT DELAYED MEASUREMENT

Opinion-Induction condition	Recall of topic and side taken	
	Yes	No
Writing		
Postinduction opinion mean	27.34	35.50
Subsequent change	+2.02 ($N = 56$)	-2.63 ($N = 8$)
Reading		
Postinduction opinion mean	28.91	26.21
Subsequent change	-3.00 ($N = 33$)	-0.51 ($N = 43$)
Overall		
Postinduction opinion mean	27.92	27.67
Subsequent change	+0.14 ($N = 89$)	-0.84 ($N = 51$)

recall and analyzed in a $2 \times 2 \times 2$ analysis of variance (Method of Induction \times Order of Measurement \times Recall).

None of the interactions involving order of measurement approached significance (the highest F ratio being 1.11); consequently, the means presented in Table 3 have been collapsed across this variable for clarity of exposition.

It is clear that the temporal superiority of the active-participation condition is not directly due to superior recall since persistence of opinion change is positively related to recall in the writing condition, but this relationship reverses to negative in the reading condition. That is, among subjects who read a persuasive message, those who are unable to recall the topic and side taken maintain more of the initially induced opinion change. This interaction effect between method of induction and recall is significant at the .05 level ($F = 5.70$). However, it must be viewed with extreme caution since partitioning the subjects on the basis of recall resulted in appreciable, although insignificant ($F = 1.87$, $df = 3/136$), differences in the immediate postinduction opinion levels. Hence, this interaction may well be due, at least in part, to regression. Comparing only those subjects who correctly recalled the topic and side taken (see the first column of Table 3), the difference between methods of induction is very pro-

nounced. Subjects who had improvised arguments showed a mean gain of 1.02 points over time compared to a mean loss of 3.18 points for their counterparts in the reading condition ($t = 3.02, p < .01$). Regression would not be an important factor in this comparison since the immediate postinduction opinion levels were comparable.

The inverse relationship between persistence of opinion change and recall obtained in the reading condition approaches significance ($t = 1.70, p < .10$) and is in general agreement with data reported by Watts and McGuire (1964) where, for subjects who had read persuasive messages, memory of the topic (but not the side taken) became inversely related to opinion change over a 6-week time interval.

Degree of Involvement Related to Method of Opinion Induction and Persistence of Opinion Change

From the work of Culbertson (1957) it was expected that subjects who had actively participated by improvising arguments would be more involved with the issue than their counterparts who had read a prepared message. Three indexes of involvement were utilized in the present study: whether the subject had subsequently discussed the issue, whether he had read anything about the issue in other sources during the intervening weeks, and, finally, how interesting he had found the task of reading or writing about the issue assigned. Each will be discussed in turn.

Subsequent discussion of the topic. Subjects rated how frequently they had discussed the topic during the intervening weeks on a 15-point scale ranging from 1 ("not at all") to 15 ("very often"). Since the resulting distribution was quite asymmetrical, with approximately half of the subjects checking "not at all," the data were dichotomized into those subjects who had and those who had not discussed the issue. As predicted, subjects who had written a persuasive argument were more likely to have discussed the topic, to some extent, during the 6-week period. Sixty-three percent of this group claimed that they had discussed the topic compared to 38% of the subjects in the reading condition. This difference in percentages is significant ($Z = 2.87,$

$p < .01$) and suggests that the temporal superiority of the improvisation condition may be due, at least in part, to the subjects' greater tendency to discuss the topic.

To examine the relationship between discussion and persistence of opinion change, the subjects' opinion scores were partitioned on the basis of whether they had or had not discussed the issue, and these means are presented in Table 4.

First, it should be noted, in this table, partitioning subjects on the basis of discussion did not result in appreciable differences in postinduction opinion levels ($F = .80, df = 3/127$). That is, there was no tendency for subjects who were more favorable at the close of the opinion-induction session to be more inclined to discuss the topic; in fact, there is a trivial difference in the opposite direction.

It can be seen in Table 4 that, irrespective of method of induction, those subjects who claim that they have discussed the topic show greater persistence of the initially induced opinion change. In fact, this group shows a mean increment of 1.06 points (horizontal margin of Table 4) which differs significantly ($t = 2.26, p < .05$) from the mean loss of 1.74 points among their counterparts who say that they have not discussed the topic. Hence, this variable would appear to account, in part,

TABLE 4
PERSISTENCE OF OPINION CHANGE AS RELATED
TO SUBSEQUENT DISCUSSION OF TOPIC

Opinion-induction condition	Subsequent discussion	
	Yes	No
Writing		
Postinduction opinion mean	27.73	29.86
Subsequent change	+2.08 ($N = 38$)	+0.41 ($N = 22$)
Reading		
Postinduction opinion mean	25.40	28.86
Subsequent change	-0.37 ($N = 27$)	-2.82 ($N = 44$)
Overall		
Postinduction opinion mean	26.76	29.14
Subsequent change	+1.06 ($N = 65$)	-1.74 ($N = 66$)

Note.—Nine subjects failed to respond to this question; consequently, the N was reduced to 131. These subjects appear to have overlooked this last page, and, in one case, it had been inadvertently omitted from the questionnaire.

for the temporal superiority of the writing condition.

That subjects in the writing condition would feel more personally involved and want to talk about the topic with others, either for purposes of self-evaluation (Festinger, 1954) or proselyting (Festinger, Riecken, & Schachter, 1956), is quite congruent with common sense. It does, however, raise a grave methodological question of whether the subsequent discussion may have restricted the independence of subjects' responses on the delayed questionnaire, thus violating the statistical model. A comparison of the standard deviations of the postinduction scores and the 6-week-delayed scores shows that they are virtually identical ($\sigma = 10.68$, $\sigma = 10.56$, respectively). Thus, from the available evidence, there is no reason to believe that independence was curtailed. Indeed, the classes utilized were not of a nature to encourage extensive interaction among class members as they met only one time a week for a 2-hour period which was devoted to formal lecture.

Subsequent reading about the topic in other sources. Since the majority (67%) of subjects stated that they had read nothing about the topic during the intervening weeks, responses to this item were also dichotomized into those who had and had not read about the topic. Thirty-eight percent of those subjects in the writing condition claimed they had read something about the topic elsewhere compared to 28% in the reading condition.⁶ While this difference is in the predicted direction, it reaches only a low level of significance ($Z = 1.14$, $p \approx .25$). The cell N s in Table 5 represent the number of subjects who did and those who did not read about the topic, and the cell means represent their changes in opinion over time.

Partitioning subjects on this basis resulted in appreciable differences ($F = 2.38$, $p < .10$, $df = 3/125$) in the immediate postinduction cell means (see Table 5), thus rendering individual cell comparisons ambiguous because of regression artifacts. However, it can be seen that the overall relationship be-

⁶ Subjects' responses to the follow-up question asking where they had read about the topic were so vague and general (i.e., books, magazines, etc.) as to be of little use and, consequently, were not analyzed.

TABLE 5
PERSISTENCE OF OPINION CHANGE AS RELATED
TO SUBSEQUENT READING ABOUT TOPIC

Opinion-induction condition	Subsequent reading	
	Yes	No
Writing		
Postinduction opinion mean	31.96	26.38
Subsequent change	+1.48 ($N = 23$)	+1.16 ($N = 37$)
Reading		
Postinduction opinion mean	24.05	28.96
Subsequent change	+1.53 ($N = 19$)	-2.86 ($N = 50$)
Overall		
Postinduction opinion mean	28.38	27.86
Subsequent change	+1.52 ($N = 42$)	-1.02 ($N = 87$)

Note.—The N is reduced to 129 because of the 9 subjects referred to in the footnote at Table 4 and 2 additional subjects who completed the last page but left this question blank.

tween subsequent reading about the topic and persistence of opinion change is positive (horizontal margin of Table 5). Subjects who say that they have read about the topic elsewhere show an increment of 1.52 points over the 6-week period, while their counterparts show a mean decline of 1.02 points. This difference in means is significant at the .05 level ($t = 2.08$) and would not be due to regression since the overall means were quite comparable (28.38 compared to 27.86). Hence, it would appear that the greater persistence of opinion change induced by improvising arguments in defense of an attitudinal position is, in part, due to the more active seeking of additional information.

Subjects' rated interest in the task. No differences were obtained between the reading and writing conditions in rated interest of the task (the means, on a 15-point scale, were 6.83 and 6.98, respectively), nor was there any relationship between interest and persistence of opinion change. When subjects were partitioned on the basis of their being above or below the median on the interest item, the opinion-change means for the two groups were virtually identical.

DISCUSSION

Subjects who improvised arguments in defense of an attitudinal position were more likely to have subsequently discussed the topic

and read about it in other sources (or at least to report that they had) than their counterparts who had read a prepared message. These activities were, in turn, positively related to retention of opinion change suggesting that the temporal superiority conferred by the active participation (writing condition) over the more passive reading condition is due, in part, to these processes.

While this relationship between method of inducing opinion change and its temporal persistence was predicted on the basis of greater involvement among subjects who had improvised arguments, the term involvement is ambiguous, and labeling the phenomenon as such does little to increase understanding of it. Hence, two questions warrant further discussion: (a) What is there about the active participation condition that leads to more frequent reading and discussion with others? (b) Given this behavior, how does it produce the differential change in opinion over time?

Perhaps the most elementary interpretation of the effects of active participation is that writing an argument is more interesting and novel than reading one prepared by a distant and impersonal source, and, therefore, these subjects are more inclined to discuss the topic or read about it afterward. However, this explanation is not tenable since there was no appreciable difference between the active and passive conditions in their rated interest in the topic.

A more compelling interpretation is based upon a consistency principle such as evoked by Lieberman (1950) as one explanation for the changes in attitude accompanying role changes, although the application is somewhat different. In the present case, subjects who wrote an essay advocating an attitudinal position committed themselves, to some degree, to that position which may have been more or less discrepant with their former belief (since the present design did not manipulate or attempt to control the magnitude of discrepancy). Of the various consistency models, Festinger's (1957) theory of cognitive dissonance would seem most applicable to this situation. According to this theory, subjects would be expected to change their attitudes so as to be consistent with the position they advocated (as the group means indicate did

happen), and, furthermore, they might be expected to seek other cognitions consonant with their behavior. That is, in order to reduce the dissonance resulting from defending a discrepant attitudinal position they might seek further supportive information to bolster the position advocated and their new belief level by selectively acquiring new information about the topic via reading and discussion. It is clear that such self-consistency pressures should be lower for those subjects who had read a message since this would not involve the same degree of personal commitment to the attitudinal position.

The results are also readily interpreted in terms of Festinger's (1954) theory of social comparison processes. After writing their persuasive essays, the subjects received no information about the quality of their product. This theory would imply that, in the absence of such feedback, subjects would want to discuss the topic with others or gain information from other sources in order to evaluate the forcefulness and logic of their own arguments. Hence, on this basis also, one would expect a higher rate of such activity for subjects who had written arguments since their counterparts in the reading condition were in a less acute state of uncertainty having passively read prepared messages underlining the most important points. (Although this is not to say that they would have had no desire for subsequent evaluation since they might wish to see what others considered the most important points to be. Indeed, the data indicate that a number of these subjects did subsequently read about or discuss the topic.)

It may also be that subjects in the active-participation condition were made dramatically aware of their own ignorance of these topics when called upon to improvise arguments. Since the issues were probably of low salience for the subjects, and the time allotted for composing arguments brief (8 minutes), it would seem unlikely that they would be able to marshal an impressive array of ideas. (The quality of the arguments attests to this point.) Hence, after writing their essays, subjects may have continued to rehearse other arguments that they might have used had they thought of them and be motivated to discuss the topic and read about it in search of

additional information as well as to ascertain the veracity of the statements they had made. Naturally, this interpretation assumes that the subject was not devastated by failure; for, if this were the case, he might want to put the entire experience out of mind because of its unpleasant aspects. This revelation of ignorance would be less acute for subjects who had read arguments since, in assimilating the information presented, they were never confronted with the question of how much they knew about the issue.

Turning to the effects that subsequent discussion of and reading about the topic had upon opinion change over the 6-week time period, the positive relationships obtained suggest considerable selectivity on the part of the subjects in their acquisition of new information. That is, if a preponderance of the new information received were to oppose their belief, one would expect a detrimental effect upon opinion change rather than the incremental effect obtained. It has long been held by researchers in the mass media that people tend to read what they already agree with (e.g., Lazarsfeld, Berelson, & Gaudet, 1948). Similarly, Festinger (1957) proposed that individuals seek consonant and avoid dissonant information. While the majority of experimental studies have failed to confirm the avoidance hypothesis (see, e.g., Bock, 1965), two recent studies (Mills, 1965; Rhine, 1966) do find support for it, and the authors offer possible explanations for the previous failures.

Although this issue of selective attention to consonant and dissonant information cannot yet be considered as resolved, it seems likely that in the present study subjects would not be exposed to strong counterattitudinal information. The topics utilized were innocuous and, with the possible exception of the "lenient treatment of juvenile delinquents" issue, were probably of low saliency for both the subjects and the population with whom they might discuss the topic. Hence, the probability of encountering strong opposition to their belief was low.

Assuming the operation of such a selection factor, the results obtained would be as expected with subjects who had discussed the topic or read about it retaining their initial opinion change or becoming even more favorable toward the position advocated with the

passage of time and the acquisition of additional favorable information. These conjectures imply that it would be unwise to generalize the present results to all topics. If the topic defended was counternorm for the population or even highly controversial, where much of the information that the subject might encounter would strongly oppose his new opinion, he might be expected to succumb to these pressures. Indeed, in this case, a reversal might be expected with information seeking adversely affecting persistence of opinion change and relative insulation of the belief promoting greater persistence.

REFERENCES

- BREHM, J. W. Comment on "Counter-norm attitudes induced by consonant versus dissonant conditions of role playing." *Journal of Experimental Research in Personality*, 1965, 1, 61-64.
- BROCK, T. C. Commitment to exposure as a determinant of information receptivity. *Journal of Personality and Social Psychology*, 1965, 2, 10-19.
- COHEN, A. R., BREHM, J. W., & FLEMING, W. H. Attitude change and justification for compliance. *Journal of Abnormal and Social Psychology*, 1958, 56, 276-278.
- COHEN, A. R., & LATANÉ, B. An experiment on choice in commitment to counterattitudinal behavior. In J. W. Brehm & A. R. Cohen (Eds.), *Explorations in cognitive dissonance*. New York: Wiley, 1962. Pp. 88-91.
- CULBERTSON, F. M. Modification of an emotionally held attitude through role playing. *Journal of Abnormal and Social Psychology*, 1957, 54, 230-233.
- DUNNETT, C. W. A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association*, 1955, 50, 1096-1121.
- ELMS, A. C., & JANIS, I. L. Counter-norm attitudes induced by consonant versus dissonant role playing. *Journal of Experimental Research in Personality*, 1965, 1, 50-60.
- FESTINGER, L. A theory of social comparison processes. *Human Relations*, 1954, 7, 117-140.
- FESTINGER, L. *A theory of cognitive dissonance*. Stanford: Stanford University Press, 1957.
- FESTINGER, L., RIECKEN, H., & SCHACHTER, S. *When prophecy fails*. Minneapolis: University of Minnesota Press, 1956.
- JANIS, I. L., & GILMORE, J. B. The influence of incentive conditions on the success of role playing in modifying attitudes. *Journal of Personality and Social Psychology*, 1965, 1, 17-27.
- JANIS, I. L., & KING, B. T. The influence of role playing on opinion change. *Journal of Abnormal and Social Psychology*, 1954, 49, 211-218.
- JANIS, I. L., & MANN, L. Effectiveness of emotional role-playing in modifying smoking habits and

- attitudes. *Journal of Experimental Research in Personality*, 1965, 1, 84-90.
- KELMAN, H. C. Attitude change as a function of response restriction. *Human Relations*, 1953, 6, 185-214.
- KELMAN, H. C., & HOVLAND, C. I. "Reinstatement" of the communicator in delayed measurement of opinion change. *Journal of Abnormal and Social Psychology*, 1953, 48, 327-335.
- KING, B. T., & JANIS, I. L. Comparison of the effectiveness of improvised versus non-improvised role playing in producing opinion changes. *Human Relations*, 1956, 9, 177-186.
- KURTZ, K. H., & HOVLAND, C. I. The effect of verbalization during observation of stimulus objects upon accuracy of recognition and recall. *Journal of Experimental Psychology*, 1953, 45, 157-164.
- LAZARSFELD, P. F., BERELSON, B., & GAUDET, H. *The people's choice*. (2nd ed.) New York: Columbia University Press, 1948.
- LEWIN, K. Group decision and social change. In E. E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. (3rd ed.) New York: Holt, Rinehart & Winston, 1958. Pp. 197-211.
- LIEBERMAN, S. The effects of changes in roles on the attitudes of role occupants. *Human Relations*, 1950, 9, 385-403.
- MILLS, J. Avoidance of dissonant information. *Journal of Personality and Social Psychology*, 1965, 2, 589-593.
- MITNICK, L. L., & MCGINNIS, E. Influencing ethnocentrism in small discussion groups through a film communication. *Journal of Abnormal and Social Psychology*, 1958, 56, 82-90.
- RABBIT, J., BREHM, J. W., & COHEN, A. R. Verbalization and reactions to cognitive dissonance. *Journal of Personality*, 1959, 27, 407-417.
- RHINE, R. J. Dissonance theory: The 1964 presidential election and curves of information selectivity. Paper read at Western Psychological Association, Long Beach, California, April 1966.
- ROSENBERG, M. J. When dissonance fails. On eliminating evaluation apprehension from attitude measurement. *Journal of Personality and Social Psychology*, 1965, 1, 28-42.
- STANLEY, J. C., & KLAUSMEIER, H. J. Opinion constancy after formal role playing. *Journal of Social Psychology*, 1957, 46, 11-18.
- WATTS, W. A., & MCGUIRE, W. J. Persistence of induced opinion change and retention of the inducing message contents. *Journal of Abnormal and Social Psychology*, 1964, 68, 233-241.

(Received March 21, 1966)

VICARIOUS EXTINCTION OF AVOIDANCE BEHAVIOR¹

ALBERT BANDURA, JOAN E. GRUSEC, AND FRANCES L. MENLOVE

Stanford University

This experiment was designed to investigate the extinction of avoidance responses through observation of modeled approach behavior directed toward a feared stimulus without any adverse consequences accruing to the model. Children who displayed fearful and avoidant behavior toward dogs were assigned to 1 of the following treatment conditions: 1 group of children participated in a series of brief modeling sessions in which they observed, within a highly positive context, a fearless peer model exhibit progressively stronger approach responses toward a dog; a 2nd group of Ss observed the same graduated modeling stimuli, but in a neutral context; a 3rd group merely observed the dog in the positive context, with the model absent; while a 4th group of Ss participated in the positive activities without any exposure to either the dog or the modeled displays. The 2 groups of children who had observed the model interact nonanxiously with the dog displayed stable and generalized reduction in avoidance behavior and differed significantly in this respect from children in the dog-exposure and the positive-context conditions. However, the positive context, which was designed to induce anxiety-competing responses, did not enhance the extinction effects produced through modeling.

Recent investigations have shown that behavioral inhibitions (Bandura, 1965a; Bandura, Ross, & Ross, 1963; Walters & Parke, 1964) and conditioned emotional responses (Bandura & Rosenthal, 1966; Berger, 1962) can be acquired by observers as a function of witnessing aversive stimuli administered to performing subjects. The present experiment was primarily designed to determine whether preexisting avoidance behavior can similarly be extinguished on a vicarious basis. The latter phenomenon requires exposing observers to modeled stimulus events in which a performing subject repeatedly exhibits approach responses toward the feared object without incurring any aversive consequences.

Some suggestive evidence that avoidance responses can be extinguished vicariously is furnished by Masserman (1943) and Jones (1924) in exploratory studies of the relative efficacy of various psychotherapeutic procedures. Masserman produced strong feeding inhibitions in cats, following which the inhibited animals observed a cage mate, that

had never been negatively conditioned, exhibit prompt approach and feeding responses. The observing subjects initially cowered at the presentation of the conditioned stimulus, but with continued exposure to their fearless companion they advanced, at first hesitantly and then more boldly, to the goal box and consumed the food. Some of the animals, however, showed little reduction in avoidance behavior despite prolonged food deprivation and numerous modeling trials. Moreover, avoidance responses reappeared in a few of the animals after the normal cat was removed, suggesting that in the latter cases the modeling stimuli served merely as temporary external inhibitors of avoidance responses. Jones (1924) similarly obtained variable results in extinguishing children's phobic responses by having them observe their peers behave in a nonanxious manner in the presence of the avoided objects.

If a person is to be influenced by modeling stimuli and the accompanying consequences, then the necessary observing responses must be elicited and maintained. In the foregoing case studies, the models responded to the most feared stimulus situation at the outset, a modeling procedure that is likely to generate high levels of emotional

¹ This research was supported by Public Health Research Grant M-5162 from the National Institute of Mental Health.

The authors are indebted to Janet Brewer, Edith Dowley, Doris Grant, and Mary Lewis for their generous assistance in various phases of this research.

arousal in observers. Under these conditions any avoidance responses designed to reduce vicariously instigated aversive stimulation, such as subjects withdrawing or looking away, would impede vicarious extinction. Therefore, the manner in which modeling stimuli are presented may be an important determinant of the course of vicarious extinction.

Results from psychotherapeutic studies (Bandura²) and experiments with infrahuman subjects (Kimble & Kendall, 1953) reveal that avoidance responses can be rapidly extinguished if subjects are exposed to a graduated series of aversive stimuli that progressively approximate the original intensity of the conditioned fear stimulus. For the above reasons it would seem advisable to conduct vicarious extinction by exposing observers to a graduated sequence of modeling activities beginning with presentations that can be easily tolerated; as observers' emotional reactions to displays of attenuated approach responses are extinguished, the fear-provoking properties of the modeled displays might be gradually increased, concluding with interactions capable of arousing relatively strong emotional responses.

If emotion-eliciting stimuli occur in association with positively reinforcing events, the former cues are likely to lose their conditioned aversive properties more rapidly (Farber, 1948) than through mere repeated nonreinforced presentation. It might therefore be supposed that vicarious extinction would likewise be hastened and more adequately controlled by presenting the modeling stimuli within a favorable context designed to evoke simultaneously competing positive responses.

The principles discussed above were applied in the present experiment, which explored the vicarious extinction of children's fearful and avoidant responses toward dogs. One group of children participated in a series of modeling sessions in which they observed a fearless peer model exhibit progressively longer, closer, and more active interactions with a dog. For these subjects, the modeled approach behavior was presented within a

highly positive context. A second group of children was presented the same modeling stimuli, but in a neutral context.

Exposure to the behavior of the model contains two important stimulus events, that is, the occurrence of approach responses without any adverse consequences to the performer, and repeated observation of the feared animal. Therefore, in order to control for the effects of exposure to the dog per se, children assigned to a third group observed the dog in the positive context but with the model absent. A fourth group of children participated in the positive activities, but they were never exposed to either the dog or the model.

In order to assess both the generality and the stability of vicarious extinction effects, the children were readministered tests for avoidance behavior toward different dogs following completion of the treatment series, and approximately 1 month later. It was predicted that children who had observed the peer model interact nonanxiously with the dog would display significantly less avoidance behavior than subjects who had no exposure to the modeling stimuli. The largest decrements were expected to occur among children in the modeling-positive context condition. It was also expected that repeated behavioral assessments and the general disinhibitory effects of participation in a series of highly positive activities might in themselves produce some decrease in avoidance behavior.

METHOD

Subjects

The subjects were 24 boys and 24 girls selected from three nursery schools. The children ranged in age from 3 to 5 years.

Pretreatment Assessment of Avoidance Behavior

As a preliminary step in the selection procedure, parents were asked to rate the magnitude of their children's fearful and avoidant behavior toward dogs. Children who received high fear ratings were administered a standardized performance test on the basis of which the final selection was made.

The strength of avoidance responses was measured by means of a graded sequence of 14 performance tasks in which the children were required to engage in increasingly intimate interactions with a dog. A female experimenter brought the children individually to the test room, which contained a brown cocker spaniel confined in a modified playpen. In the initial tasks the children were asked, in

² A. Bandura, "Principles of Behavioral Modification," unpublished manuscript, Stanford University, 1966.

the following order, to walk up to the playpen and look down at the dog, to touch her fur, and to pet her. Following the assessment of avoidance responses to the dog in the protective enclosure, the children were instructed to open a hinged door on the side of the playpen, to walk the dog on a leash to a throw rug, to remove the leash, and to turn the dog over and scratch her stomach. Although a number of the subjects were unable to perform all of the latter tasks, they were nevertheless administered the remaining test items to avoid any assumption of a perfectly ordered scale for all cases. In subsequent items the children were asked to remain alone in the room with the animal and to feed her dog biscuits. The final and most difficult set of tasks required the children to climb into the playpen with the dog, to pet her, to scratch her stomach, and to remain alone in the room with the dog under the exceedingly confining and fear-provoking conditions.

The strength of the children's avoidant tendencies was reflected not only in the items completed, but also in the degree of vacillation, reluctance, and fearfulness that preceded and accompanied each approach response. Consequently, children were credited 2 points if they executed a given task either spontaneously or willingly, and 1 point when they carried out the task minimally after considerable hesitancy and reluctance. Thus, for example, children who promptly stroked the dog's fur repeatedly when requested to do so received 2 points, whereas subjects who held back but then touched the dog's fur briefly obtained 1 point. In the item requiring the children to remain alone in the room with the dog, they received 2 points if they approached the animal and played with her, and 1 point if they were willing to remain in the room but avoided any contact with the dog. Similarly, in the feeding situation children were credited 2 points if they fed the dog by hand, but a single point if they tossed the biscuits on the floor and thereby avoided close contact with the animal. The maximum approach score that a subject could attain was 28 points.

On the basis of the pretreatment assessment, the children in each nursery school were grouped into three levels of avoidance behavior, with the corresponding scores ranging from 0 to 7, 8 to 17, and 18 to 20 points. There were approximately the same number of children, equally divided between boys and girls, at each of the three avoidance levels. The subjects from each of these groups were then assigned randomly to one of four conditions.

Treatment Conditions

Children who participated in the *modeling-positive context* condition observed a fearless peer model display approach responses toward a cocker spaniel within the context of a highly enjoyable party atmosphere.

There were eight 10-minute treatment sessions conducted on 4 consecutive days. Each session,

which was attended by a group of four children, commenced with a jovial party. The children were furnished brightly colored hats, cookie treats, and given small prizes. In addition, the experimenter read stories, blew large plastic balloons for the children to play with, and engaged in other party activities designed to produce strong positive affective responses.

After the party was well under way, a second experimenter entered the room carrying the dog, followed by a 4-year-old male model who was unknown to most of the children. The dog was placed in a playpen located across the room from a large table at which the children were seated. The model, who had been chosen because of his complete lack of fear of dogs, then performed prearranged sequences of interactions with the dog for approximately 3 minutes during each session. One boy served as the model for children drawn from two of the nursery schools, and a second boy functioned in the same role at the third school.

The fear-provoking properties of the modeled displays were gradually increased from session to session by varying simultaneously the physical restraints on the dog, the directness and intimacy of the modeled approach responses, and the duration of interaction between the model and his canine companion. Initially, the experimenter carried the dog into the room and confined her to the playpen, and the model's behavior was limited to friendly verbal responses ("Hi, Chloe") and occasional petting. During the following three sessions the dog remained confined to the playpen, but the model exhibited progressively longer and more active interactions in the form of petting the dog with his hands and feet, and feeding her wieners and milk from a baby bottle. Beginning with the fifth session, the dog was walked into the room on a leash, and the modeled tasks were mainly performed outside the playpen. For example, in addition to repeating the feeding routines, the model walked the dog around the room, petted her, and scratched her stomach while the leash was removed. In the last two sessions the model climbed into the playpen with the dog where he petted her, hugged her, and fed her wieners and milk from the baby bottle.

It would have been of interest to compare the relative efficacy of the graduated modeling technique with bold displays of approach behavior from the outset. However, pretest findings showed that when modeled displays are too fear provoking, children actively avoid looking at the performances and are reluctant to participate in subsequent sessions. The latter approach would therefore require additional procedures designed to maintain strong attending behavior to highly aversive modeling stimuli.

Children assigned to the *modeling-neutral context* condition observed the same sequence of approach responses performed by the same peer model except that the parties were omitted. In each of the eight sessions the subjects were merely seated at the table and observed the modeled performances.

In order to control for the influence of repeated exposure to the positive atmosphere and to the dog per se, children in the *exposure-positive context* group attended the series of parties in the presence of the dog with the model absent. As in the two modeling conditions, the dog was introduced into the room in the same manner for the identical length of time; similarly, the dog was confined in the playpen during the first four sessions and placed on a leash outside the enclosure in the remaining sessions.

Children in the *positive-context* group participated in the parties, but they were never exposed to either the dog or the model. The main purpose of this condition was to determine whether the mere presence of a dog had an adverse or a beneficial effect on the children. Like the third condition, it also provided a control for the possible therapeutic effects of positive experiences and increased familiarity with amiable experimenters, which may be particularly influential in reducing inhibitions in very young children. In addition, repeated behavioral assessments in which subjects perform a graded series of approach responses toward a feared object without any aversive consequences would be expected to produce some direct extinction of avoidance behavior. The inclusion of the latter two control groups thus makes it possible to evaluate the changes effected by exposure to modeling stimuli over and above those resulting from general disinhibition, direct extinction, and repeated observation of the feared object.

Posttreatment Assessment of Avoidance Behavior

On the day following completion of the treatment series, the children were readministered the performance test consisting of the graded sequence of interaction tasks with the dog. In order to determine the generality of vicarious extinction effects, half the children in each of the four groups were tested initially with the experimental animal and then with an unfamiliar dog; the remaining children were presented with the two dogs in the reverse order.³ The testing sessions were separated by an interval of 1½ hours so as to minimize any transfer of emotional reactions generated by one animal to the other.

The unfamiliar animal was a white mongrel, predominantly terrier, and of approximately the same size and activity level as the cocker spaniel. Two groups of 15 children, drawn from the same nursery-school population, were tested with either the mongrel or the spaniel in order to determine the aversiveness of the two animals. The mean approach scores with the spaniel ($M = 16.47$) and the mongrel ($M = 15.80$) were virtually identical ($t = .21$).

³ The authors are especially indebted to Chloe and Jenny for their invaluable and steadfast assistance with a task that, at times, must have been most perplexing to them.

Follow-Up Assessment

A follow-up evaluation was conducted approximately 1 month after the posttreatment assessment in order to determine the stability of modeling-induced changes in approach behavior. The children's responses were tested with the same performance tasks toward both animals, presented in the identical order.

After the experiment was completed, the children were told that, while most dogs are friendly, before petting an unfamiliar dog they should ask the owner. This precautionary instruction was designed to reduce indiscriminate approach behavior by children who were in the modeling conditions toward strange dogs which they would undoubtedly encounter.

Measurement Procedure

The same female experimenter administered the pretreatment, posttreatment, and follow-up behavioral tests. To prevent any possible bias, the experimenter was given minimal information about the details of the study and had no knowledge of the conditions to which the children were assigned. The treatment and assessment procedures were further separated by the use of different rooms for each activity.

In order to provide an estimate of interscorer reliability, the performances of 25% of the children, randomly selected from pretreatment, posttreatment, and follow-up phases of the experiment, were scored simultaneously but independently by another rater who observed the test sessions through a one-way mirror from an adjoining observation room. The two raters were in perfect agreement on 97% of the specific approach responses that were scored.

A dog's activity level may partly determine the degree of fear and avoidance exhibited by the children; conversely, timorous or unrestrained approach responses might differentially affect the animals' reactivity. Therefore, during the administration of each test item, the animals' behavior was rated as either passive, moderately active, or vigorous. The raters were in perfect agreement in categorizing the dogs' activity levels on 81% of the performance tests.

Changes in children's approach-response scores across the different phases of the experiment, and the number of subjects in each treatment condition who were able to carry out the terminal performance task served as the dependent measures.

RESULTS

The percentages of test items in which the animals behaved in a passive, moderately active, or vigorous manner were 55, 43, and 2, respectively, for the model-positive context group; 53, 44, and 2 for children in the model-neutral context condition; 52, 45, and

3 for the exposure-positive context group; and 57, 41, and 2 for the positive-context subjects. Thus, the test animals did not differ in their behavior during the administration of performance tasks to children in the various treatment conditions.

Approach Responses

Table 1 presents the mean increases in approach behavior achieved by children in each of the treatment conditions in different phases of the experiment with each of the test animals.

The children's approach responses toward the two dogs did not differ either in the post-treatment assessment ($t = 1.35$) or in the follow-up phase ($t = .91$) of the study. Nor were there any significant effects ($t = 1.68$) due to the order in which the test animals were presented following completion of the treatment series. A t -test analysis also disclosed no significant change ($t = 1.50$) in mean approach scores between measurements conducted in the posttreatment and the follow-up phases of the experiment. Moreover, analysis of variance of the posttreatment scores revealed no significant Treatment \times Dogs ($F = 2.15$) or Treatment \times Order ($F = .30$) interaction effects. The data were therefore combined across phases and test animals in evaluating the major hypotheses.

An analysis of covariance, in which adjustments were made for differences in initial level of avoidance, was computed for mean approach responses performed by children in the various groups. The results reveal that the treatment conditions had a highly signifi-

cant effect on the children's behavior ($F = 5.09$, $p < .01$). Tests of the differences between the various pairs of treatments indicate that subjects in the modeling-positive context condition displayed significantly more approach behavior than subjects in either the exposure ($F = 9.32$, $p < .01$) or the positive-context ($F = 8.96$, $p < .01$) groups. Similarly, children who had observed the model within the neutral setting exceeded both the exposure ($F = 6.57$, $p < .05$) and positive-context groups ($F = 4.91$, $p < .05$) in approach behavior. However, the data yielded no significant differences between either the two modeling conditions ($F = .04$) or the two control groups ($F = .76$).

Within-Group Analysis of Approach Responses

The approach scores obtained by the different groups of children in preexperimental and subsequent tests are summarized graphically in Figure 1. Within-group analyses of changes between initial performance and mean level of approach behavior following treatment disclose significant increases in approach behavior for children in the modeling-positive context group ($t = 7.71$, $p < .001$) and for those who observed the modeling performance within the neutral setting ($t = 5.80$, $p < .001$). Although the positive-context group showed an increment in approach behavior ($t = 5.78$, $p < .001$), children who were merely exposed to the dog in the positive context achieved a small, but nonsignificant ($t = 1.98$), reduction in avoidance responses.

TABLE 1
MEAN INCREASES IN APPROACH RESPONSES AS A FUNCTION OF TREATMENT
CONDITIONS, ASSESSMENT PHASES, AND TEST ANIMALS

Phases	Treatment conditions			
	Modeling—positive context	Modeling—neutral context	Exposure—positive context	Positive context
Posttreatment				
Spaniel	10.83	9.83	2.67	6.08
Mongrel	5.83	10.25	3.17	4.17
Follow-Up				
Spaniel	10.83	9.33	4.67	5.83
Mongrel	12.59	9.67	4.75	6.67
Combined data	10.02	9.77	3.81	5.69

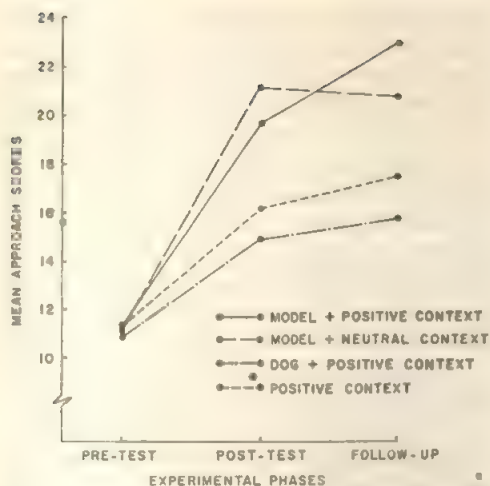


FIG. 1. Mean approach scores achieved by children in each of the treatment conditions on the three different periods of assessment.

Terminal Performances

Another measure of the efficacy of modeling procedures is provided by comparisons of the number of children in each condition who performed the terminal approach behavior at least once during the posttreatment assessment. Since the frequencies within the two modeling conditions did not differ, and the two control groups were essentially the same, the data for each of the two sets of subgroups were combined. The findings show that 67% of the children in the modeling treatment were able to remain alone in the room confined with the dog in the playpen, whereas the corresponding figure for the control subjects is 33%. The χ^2 value for these data is 4.08, which is significant beyond the .025 level.

Within the control groups, the terminal performances were attained primarily by subjects who initially showed the weakest level of avoidance behavior. The differences between the two groups are, therefore, even more pronounced if the analysis is conducted on the subjects whose pretreatment performances reflected extreme or moderately high levels of avoidance behavior. Of the most avoidant subjects in each of the two pooled groups, 55% of the children in the modeling conditions were able to perform the terminal approach behavior following the experimental sessions, while only 13% of the control sub-

jects successfully completed the final task. The one-tailed probability for the obtained $\chi^2 = 4.74$ is slightly below the .01 level of significance.

The relative superiority of the modeling groups is also evident in the follow-up phase of the experiment. Based on the stringent criterion in which the most fearful task is successfully performed with *both* animals, a significantly larger number of children in the modeling conditions (42%) than in the control groups (12%) exhibited generalized extinction ($\chi^2 = 4.22$, $p < .025$). Moreover, not a single control subject from the two highest levels of avoidance behavior was able to remain alone in the room confined in the playpen with each of the dogs, whereas 33% of the most avoidant children in the modeling conditions successfully passed both terminal approach tasks ($\chi^2 = 4.02$, $p < .025$).

DISCUSSION

The findings of the present experiment provide considerable evidence that avoidance responses can be successfully extinguished on a vicarious basis. This is shown in the fact that children who experienced a gradual exposure to progressively more fearful modeled responses displayed extensive and stable reduction in avoidance behavior. Moreover, most of these subjects were able to engage in extremely intimate and potentially fearful interactions with test animals following the treatment series. The considerable degree of generalization of extinction effects obtained to the unfamiliar dog is most likely due to similar stimulus properties of the test animals. Under conditions where observers' avoidance responses are extinguished to a single animal, one would expect a progressive decrement in approach behavior toward animals of increasing size and fearfulness.

The prediction that vicarious extinction would be augmented by presenting the modeling stimuli within a highly positive context was not confirmed, although subjects in the latter condition differed more significantly from the controls than children who observed approach behavior under neutral conditions. It is entirely possible that a different temporal ordering of emotion-provoking modeling stimuli and events designed to induce anxi-

ety-inhibiting responses would facilitate the vicarious extinction process. On the basis of evidence from conditioning studies (Melvin & Brown, 1964) the optimal treatment procedure might require repeated observational trials, in each of which aversive modeling stimuli are immediately followed by positively reinforcing experiences for the observers. These temporal prerequisites depend upon the abrupt presentation and termination of the two sets of stimulus events that cannot be readily achieved with live demonstrations. It would be possible, however, to study the effects of systematic variations in the temporal spacing of critical variables if modeling stimuli were presented pictorially. Apart from issues of economy and control, if pictorial stimulus material proved equally as efficacious as live modeling, then skillfully designed therapeutic films could be developed and employed in preventive programs for eliminating common fears and anxieties before they become well established and widely generalized.

Although children in both the exposure and the positive-context groups showed some increment in approach behavior, only the changes in the latter group were of statistically significant magnitude. Apparently the mere presence of a dog had some mild negative consequences that counteracted the facilitative effects resulting from highly rewarding interactions with amiable experimenters, increased familiarity with the person conducting the numerous tests of avoidance behavior, and any inevitable direct extinction produced by the repeated performance of some approach responses toward the test animals without any adverse consequences. As might be expected, the general disinhibitory effects arising from these multiple sources occurred only in the early phase of the experiment, and no significant increases in approach behavior appeared between the post-treatment and follow-up assessments.

The data obtained in this experiment demonstrate that the fearless behavior of a model can substantially reduce avoidance responses in observers, but the findings do not establish the nature of the mechanism by which vicarious extinction occurs. There are several possible explanations of vicariously produced

effects (Bandura, 1965b; Kanfer, 1965). One interpretation is in terms of the informative value of modeling stimuli. That is, the repeated evocation of approach responses without any adverse consequences to another person undoubtedly conveys information to the observer about the probable outcomes of close interactions with dogs. In the present study, however, an attempt was made to minimize the contribution of purely cognitive factors by informing children in all groups beforehand that the test animals were harmless.

The nonoccurrence of anticipated aversive consequences to a model accompanied by positive affective reactions on his part can also extinguish in observers previously established emotional responses that are vicariously aroused by the modeled displays (Bandura & Rosenthal, 1966). It is therefore possible that reduction in avoidance behavior is partly mediated by the elimination of conditioned emotionality.

Further research is needed to separate the relative contribution of cognitive, emotional, and other factors governing vicarious processes. It would also be of interest to study the effects upon vicarious extinction exercised by such variables as number of modeling trials, distribution of extinction sessions, mode of model presentation, and variations in the characteristics of the models and the feared stimuli. For example, with extensive sampling in the modeled displays of both girls and boys exhibiting approach responses to dogs ranging from diminutive breeds to larger specimens, it may be possible to achieve widely generalized extinction effects. Once approach behaviors have been restored through modeling, their maintenance and further generalization can be effectively controlled by response-contingent reinforcement administered directly to the subject. The combined use of modeling and reinforcement procedures may thus serve as a highly efficacious mode of therapy for eliminating severe behavioral inhibitions.

REFERENCES

- BANDURA, A. Influence of models' reinforcement contingencies on the acquisition of imitative responses. *Journal of Personality and Social Psychology*, 1965, 1, 589-595. (a)
- BANDURA, A. Vicarious processes: A case of no-trial learning. In L. Berkowitz (Ed.), *Advances in ex-*

- perimental social psychology* Vol 2 New York: Academic Press, 1965. Pp. 1-55. (b)
- HANDURA, A., & ROSENTHAL, T. L. Vicarious classical conditioning as a function of arousal level. *Journal of Personality and Social Psychology*, 1966, 3, 54-62.
- HANDURA, A., ROSS, D., & ROSS, S. A. Vicarious reinforcement and imitative learning. *Journal of Abnormal and Social Psychology*, 1963, 67, 601-607.
- BEGER, S. M. Conditioning through vicarious instillation. *Psychological Review*, 1962, 69, 450-466.
- FARRER, I. E. Response fixation under anxiety and non-anxiety conditions. *Journal of Experimental Psychology*, 1948, 38, 111-131.
- JONES, M. C. The elimination of children's fears. *Journal of Experimental Psychology*, 1924, 7, 383-390.
- KANFER, F. H. Vicarious human reinforcement: A glimpse into the black box. In L. Krasner & L. P. Ullmann (Eds.), *Research in behavior modification*. New York: Holt, Rinehart & Winston, 1966. Pp. 244-267.
- KIMBLE, G. A., & KENDLER, J. W. Jr. A comparison of two methods of producing experimental extinction. *Journal of Experimental Psychology*, 1952, 45, 87-96.
- MASSERMAN, J. H. *Behavior and nervous system*. Chicago: University of Chicago Press, 1963.
- MELVIN, K. B., & BROWN, J. S. Neutralization of an aversive light stimulus as a function of number of paired presentations with food. *Journal of Comparative and Physiological Psychology*, 1964, 58, 350-353.
- WATERS, R. H., & PARKY, R. D. Influence of response consequences to a social model on resistance to deviation. *Journal of Experimental Child Psychology*, 1964, 1, 269-280.

(Received January 13, 1966)

WAITING FOR REWARDS AND PUNISHMENTS: EFFECTS OF TIME AND PROBABILITY ON CHOICE¹

WALTER MISCHEL AND JOAN GRUSEC²

Stanford University

Children made real choices between immediate smaller and delayed larger rewards, and between immediate smaller and delayed larger punishments. For $\frac{1}{2}$ of the Ss delay time for the larger rewards and punishments was varied (1 day, 1 wk., 1 mo.), and for the remaining children probability for the occurrence of the larger outcomes was varied ($P = .1, .5, 1.0$). $\frac{1}{2}$ of the Ss made reward choices before punishment choices, while this sequence was reversed for the others. As expected, increasing probability led to more delayed reward choices and more immediate punishment choices. Surprisingly, although more immediate reward choices were made as delay time increased, punishment choices were independent of the amount of delay time. Moreover, voluntary delay for rewards and punishments was essentially unrelated, suggesting at least partially different determinants for these 2 types of choice behavior. Finally, Ss who made reward choices before punishment choices were more willing to accept immediate punishments.

The effects of externally imposed delay of reinforcement on learning and performance have long been the focus of extensive laboratory research (e.g., Mowrer, 1960; Renner, 1964). In contrast, voluntary delay of reward, although recognized widely as important (Freud, 1959; Rapaport, 1950; Singer, 1955), has been investigated much less thoroughly. Self-imposed delay of reinforcement is a critical aspect of many complex human activities, and an adequate personality theory must include a conceptualization of the determinants of this behavior.

During the last decade there have been a number of empirical inquiries into the antecedents and correlates of the willingness to defer immediate, less valued rewards for the sake of more valuable but temporally deferred outcomes. These studies typically have employed a paradigm in which subjects made real choices between immediately available but smaller positive outcomes as opposed to delayed but larger rewards (e.g., Mahrer, 1956; Mischel, 1958, 1961c). For example, children in these studies chose between a

little candy bar now and a larger one whose attainment required waiting a week. Results from such procedures provide evidence that delay responses are relatively stable, tend to increase with age, and are systematically related to other theoretically relevant variables usually subsumed under "ego strength" constructs (e.g., Mischel, 1961a, 1961b, 1966). Positive relationships have been obtained between willingness to wait for more valued rewards and measures of self-control, achievement motivation, certain rearing conditions, and a variety of intellectual and cognitive indexes (Mischel, 1966). In addition, some of the social learning variables which determine such choice behavior have been investigated (Bandura & Mischel, 1965; Mischel & Staub, 1965).

The present conceptualization of delay behavior is based on a social learning theory in which each choice is a function of both the subjective expectancy that it will lead to particular consequences in a given situation and the subjective value of those consequences (Rotter, 1954). Thus, whether an individual chooses an immediate smaller reward or a delayed larger reward in a given situation is a function of the relative strengths of the expectancy and reward value associated with each choice. Some support for this comes from studies by Mahrer (1956) and by H. Mischel (1963), in which increases in pref-

¹ This investigation was supported by Research Grant M-6830 from the National Institutes of Health, United States Public Health Service. The authors are indebted to the officials and teachers of the Palo Alto and Mountain View schools who cooperated in this study and to Tele Bjork who assisted in collecting the data.

² Now at University of Waterloo.

ferences for delayed rewards resulted from strengthening children's experimentally manipulated trust in the promise maker. Likewise, decreasing the length of the delay interval has been shown to increase the frequency with which delayed positive rewards are chosen (Mischel & Metzner, 1962), presumably by increasing the expectancy that the delayed reward will actually materialize. Moreover, when the expectancy for obtaining contingent larger rewards is strengthened by prior success experiences, individuals become more willing to work for them (Mischel & Staub, 1965).

Although the willingness to wait for rewards has been studied to a considerable extent, another aspect of voluntary delay behavior has been ignored entirely. In addition to choosing between positive outcomes which vary in their temporal availability and reward value, humans often must choose between noxious alternatives. Consider, for example, the common choice of undergoing minor discomfort immediately, as in the dentist's office, or risking even more negative but delayed consequences. The determinants of this type of choice, in which negative outcomes of varying intensity and temporal distance are pitted against each other, remain unknown and are a main concern of this paper.

The present study had several aims. In our conceptualization of delay behavior the expectancy that the delayed larger outcome will occur is a main determinant of choice. Therefore, the probability of occurrence of the delayed outcome was varied systematically, for choices having both negative and positive valence. It also seems plausible that variations in temporal interval before the occurrence of the delayed outcome should affect choice by influencing subjective probability. Although data showing decrements in choice preference for delayed rewards as a function of the length of the delay interval are available (Mischel & Metzner, 1962), there is no information regarding the effects of time on the willingness to accept smaller immediate punishments as opposed to waiting for delayed larger negative outcomes. Consequently, delay time for larger positive and negative outcomes was used as the second independent

variable in this experiment. Another problem investigated was the relationship between delay behavior for immediate and delayed positive outcomes as opposed to that for immediate and delayed negative outcomes.

To study these questions a design was used in which children made real choices between immediate smaller or delayed larger positive outcomes (rewards) and between immediate smaller or delayed larger negative outcomes (punishments). In half the groups the delay interval was constant, while the probability for the delayed larger rewards and punishments was varied (from $P = .1$ to $.5$ to 1.0 , and the reverse sequence). In the remaining groups probability was constant (at $P = 1$), while the delay interval for the larger outcomes varied (from 1 day to 1 week to 1 month, and the reverse sequence).

This study also explored the effects of an initial rewarding or positive experience, as opposed to no experience, on subsequent choices between immediate and delayed punishments. Likewise, the design tested the effects of an initial negative experience, as opposed to no experience, on subsequent choices between immediate and delayed rewards. For these purposes, half the children had the presumably rewarding experience of choosing between rewards before making their choice between punishments, while this sequence was reversed for the other half.

In accord with the expectancy-value formulation of delay behavior, it was anticipated that choice preference for delayed larger rewards would increase as a function of the probability of their attainment and decrease as a function of the temporal delay period. The opposite prediction was made about choice preference for delayed larger punishments. Namely, choices for immediate smaller punishments should increase as the probability of the occurrence of the delayed larger negative alternative becomes greater and as the temporal delay before the occurrence of the delayed larger punishment decreases.

No predictions were made concerning the relationship between delay behavior for rewards and punishments nor for the effects of first choosing between rewards or punishments.

Method

Subjects

The subjects were 45 boys and 45 girls, of heterogeneous backgrounds, recruited from the fourth and fifth grades of two elementary schools in the Palo Alto-Mountain View area.

Design

Table 1 summarizes the experimental design. All subjects made choices between smaller immediate and larger delayed rewards and between smaller immediate and larger delayed punishments. Specifically, 15 boys and 15 girls, half randomly selected from one school and half from the second, were assigned to each of four main experimental conditions. In Conditions I and III (see Table 1) choices between rewards preceded those between punishments, while the reverse choice sequence occurred in Conditions II and IV. Within each choice sequence, for both the subjects (I and II), the probability for obtaining the delayed outcome was stated as certain ($P = 1.0$), while the delay interval for both delayed rewards and punishments was varied (1 day, 1 week, or 1 month). For the remaining subjects (III and IV) the delay interval was constant at 1 week, while the stated probability for both the delayed rewards and punishments was varied (from $P = 1$ to .5 to 1.0). When delay time was varied, half the subjects chose in a sequence in which delay time increased systematically and half in a sequence in which it decreased systematically. Similarly, when probability was varied half chose in an increasing sequence and half in a decreasing sequence. A total of eight testing sessions were conducted, two in each condition and half in each school to obtain the counterbalancing of school and of delay time and probability sequences, as shown in Table 1.

Procedure

The 15 subjects in each testing session were brought to the school auditorium by the experi-

mentary female assistant. The experimenter explained to the group that she was interested in the kinds of choices children make and would be asking them to make two choices between different things.

In Conditions I and III reward choices preceded punishment choices. Specifically, the experimenter gave an example of a positive choice: "You can win \$5.00 in 1 week" and explained to the children that they would have two such choices. She emphasized the importance of their choosing carefully and rationally since they would actually receive one of these choices, although they would not find out which it was until the end of the experimental session. The choices were given individually, leaving children about 30 seconds at each set of paired objects and the associated time interval. Back rewards in each pair were displayed by the experimenter who also stated the delay interval and probability for the delayed outcome, and the children recorded their choices in their booklets before the experimenter proceeded to the next choice pair.

The two positive choices consisted of three pairs of items displayed in the same sequence for each of the three delay intervals (Conditions I or the three probability levels (Conditions III). The three first pairs were \$5.00 today or \$5.00 later, two pairs of chewing gum today or three pairs of chewing gum later, and a small address book today or a large address book later.

For each positive choice subjects were confronted with one small reward, described as something better paired with a large reward, described as something in 1 day, 1 week, or 1 month, depending on the specific delay contingency. In presenting probability levels the experimenter emphasized in all conditions that if the smaller reward were chosen subjects would definitely receive it "today." Moreover, when the probability for the delayed reward was 1.0 the experimenter said that she would "definitely be back" at the indicated time with the larger reward if it were chosen. When the probability was .5, she stated that the chances of actually getting the larger reward were "50-50. You have an equal chance of either getting it or not getting it." For a probability of .1 the experimenter announced, "the chances of getting the larger reward are very small although it is possible—about 1 chance out of 10."

The punishment choices were administered immediately after the reward choices in Conditions I and III. The experimenter told subjects that they would have to make some more choices, this time between unpleasant things. She gave an example of a negative choice: "memorize perfectly 10 lines of dull poetry now or memorize perfectly 50 lines of dull poetry in 1 week" and explained that they would have 12 such choices.² Just as with the positive choices, the experimenter emphasized to the

TABLE 1

SUMMARY OF EXPERIMENTAL DESIGN

Condition	Choice sequence	Choice sequence	Delay time (in week)	Probability
I	a	Rewards-Punishments	1, 1-10 40, 7-1	1.0
II	b	Punishments-Rewards	1, 1-10 40, 7-1	1.0
III	a	Rewards-Punishments	7	1.0, .5, 1 1, .5, 1.0
IV	b	Punishments-Rewards	7	1.0, .5, 1 1, .5, 1.0

Note. In conditions I and II test the effects of choice sequence, delay time, and delay sequence. In conditions III and IV test the effects of choice sequence, probability, and probability sequence. Item letter in the school column represents a separate testing session.

² There were more punishment choices than reward choices because a secondary purpose of this study was to develop an item pool of punishment choices, while extensive prior work has already made available extensive measures for observing reward preferences.

emphasize the importance of choosing carefully what they would actually have in the event the choice is one choice pair, although they would not know what it was until the very end. The subjects were then directed to turn in their handwriting samples which had not been allowed to be perused. These pages contained brief descriptions of each member of the pair and were further examined by the experimenter.

The 11 punishment choices consisted of two items assigned to the same response for each of the three delay intervals (Condition I) or for each of the three probability levels (Condition III). The four items involved: IV response with an unpleasant teacher now or 12 months with an unpleasant teacher now, two very difficult feeling tasks to be performed by the now or six very difficult feeling tasks to be performed at three later dates, a plate of food having food now or a plate of absolutely non-functuating food later, and something positively pleasant at that point now or something positively pleasant at that point later.² In the actual situation, either one of these choices was signalled more randomly. For example, in referring to the new teacher, the experimenter said:

We have a new teacher coming in to work with you. Let me warn you, she has a very peculiar person. She is rather cranky, she is queer, larger

² In order to assess the effectiveness of the selected punishment choices, subjects at one school were asked to rate the subjective unpleasantness of these items after they had completed all their reward and punishment choices. They were presented with each pair of immediate smaller and delayed larger punishments, without their associated times, and asked to indicate which in each pair was most unpleasant or to indicate if they felt that neither was unpleasant. No subject rated the immediate punishment as more unpleasant, and no more than 10% rated any one of the four pairs as not unpleasant.

with her, just so, and she really gets on my nerves. Especially when she sees that you are nervous and that's all right, isn't it?

Again, after describing the items, or the same sentence using the last sentence just said, the experimenter further emphasized that the consequences held for two small now with some response, warning to be careful to not feel better and to not think of it, or otherwise similar saying. These warnings and probabilities were presented in the same form described for reward items.

In Conditions II and IV the negative items were administered first, followed by the positive choices. In the procedure was otherwise the same as described for Conditions I and III. It should be noted that since the old subjects knew about the second set of choices they would have to make sure they had remembered the first set. As Table 1 indicates, both the subjects chose in a manner in which later was increased, while the first had shown to a response in which it decreased. Again, when probabilities varied, that choice in an increasing manner while the remaining choice in a decreasing manner.

At the end of the administration, subjects were told which reward choice they would really get and which punishment they would really receive at the specified time. They were also informed that they would not have to do any of the unpleasant things. In addition, they were requested not to discuss any part of the experimental procedure with their fellow students, and the necessity of this was stressed. Informal postexperimental inquiries indicated that the subjects were doing about the experiment before participating.

RESULTS

Inspection³ of the data revealed no systematic differences in the choices made by boys and girls. Thus their data were combined for

TABLE 2
MEAN DELAYED REWARD AND IMMEDIATE PUNISHMENT CHOICES IN ALL CONDITIONS

Delay time in days	M	M	Probability	M	M	Choice sequence	M ⁴	M ⁵
	Reward ⁶	Punish ⁷		Reward ⁶	Punish ⁷		Reward ⁶	Punish ⁷
1	2.25	3.25	1	.55	2.05	Reward 1, punish ⁷	3.6	10.3
7	1.75	3.15	5	1.05	2.50	Punish 1, reward ⁶	5.0	8.3
30	1.30	3.05	10	1.90	3.40			
Delay Sequence	M	M	Probability Sequence	M	M	Reward 1 ⁶ - 10 ⁷ Punish 1, reward ⁶	3.1	8.9
	Reward ⁶	Punish ⁷		Reward ⁶	Punish ⁷			
1 7 30	4.2	8.8	1 5 10	3.7	8.4		4.0	7.0
30 7 1	6.4	10.1	10 5 1	3.4	7.5			

⁶ Reward of 3.

⁷ Punishment of 4.

⁸ Reward of 9.

⁹ Punishment of 12.

¹⁰ P = 1, delay time varied.

¹¹ Delay time = 7 days, probability varied.

TABLE 3

t VALUES FOR COMPARISONS OF PAIRS OF MEANS
BETWEEN CONDITIONS

Choices	Delay time (in days)		
	1 vs. 7	7 vs. 30	1 vs. 30
Rewards	4.61**	4.13**	8.72**
Punishments	NS	NS	NS
	Probability		
	.1 vs. .5	.5 vs. 1.0	.1 vs. 1.0
Rewards	3.85**	6.54**	10.38**
Punishments	2.85*	5.70**	8.54**

* $p < .01$.
** $p < .001$.

the analyses. Table 2 presents the mean number of choices of delayed larger rewards and of immediate smaller punishments made in all conditions.

Effects of Delay Time

Conditions I and II were used to assess the effect of delay time on choices and of the possible influence of the sequence in which these time intervals were presented. An analysis of variance was performed on the effects of the three delay intervals and the two delay sequences on reward choices. Delay time strongly affected preference for delayed larger rewards ($F = 19.28$, $df = 2, 92$, $p < .001$). Delay sequence also affected reward choices ($F = 17.13$, $df = 1, 46$, $p < .001$) with the sequence 30 days to 1 day leading to more delayed reward choices than the sequence 1 day to 30 days (see Table 2). There was no interaction between delay time and delay sequence ($F = .58$). Comparisons by t tests between the three delay intervals yielded highly significant differences in the expected direction, with subjects choosing fewer delayed rewards as the delay interval increased (see Table 3).

The same analyses were performed on the data for punishment choices. Contrary to prediction, delay time did not affect choice of immediate or delayed punishments ($F = 1.03$, $df = 2, 92$). Likewise, there was no significant effect of delay sequence ($F = 2.65$, $df = 1, 46$) and no interaction ($F = .22$) between delay time and delay sequence.

Effects of Probability

To assess the effect of probability on delayed reward choices, and any effect of the sequence in which probability levels were presented, the data from Conditions III and IV were analyzed. Analysis of variance for the effects of the three probability levels and the two probability sequences on reward choices showed that probability level significantly affected delayed reward choices ($F = 27.75$, $df = 2, 92$, $p < .005$). The sequence in which the probabilities were presented had no effect ($F = .14$, $df = 1, 46$), and there was no significant interaction between probability level and probability sequence ($F = 1.46$). As predicted, comparisons by t tests between the three probability levels were significant, with fewer delayed rewards chosen as the probability for obtaining the delayed outcome decreased (see Table 3).

The same analyses were performed on the data for negative choices. As expected, probability level had a significant effect on immediate negative choices ($F = 19.81$, $df = 2, 92$, $p < .005$). Probability sequence had no effect ($F = .96$), and again there was no interaction between probability level and probability sequence ($F = 1.38$). Comparisons by t tests of immediate negative choices at the three probability levels (Table 3) show that these choices increased as the probability associated with the delayed outcome increased.

Effects of the Sequence of Reward and Punishment Choices

Two analyses were done to compare the delay of reward behavior of subjects who chose initially between punishments and those who began immediately with reward choices. The first was between subjects of this sequence in Conditions I and II, in which delay time also had been manipulated, while the second tested the effects of this sequence by comparing subjects in Conditions III and IV, where probability level also had been manipulated. These analyses of variance were for the effects of choice sequence and delay time and the effects of choice sequence and probability, respectively. F s of 1.12 in Conditions I and II and 1.93 in Conditions III

and IV were obtained for the main effects of choice sequence. The interactions between choice sequence and delay time and between choice sequence and probability were not significant ($F_s = .78, .41$, respectively). Thus choice sequence did not affect preferences for immediate or delayed rewards. The main effects of delay time and probability have been discussed earlier.

Two more analyses of variance, between the same groups of subjects, were done to assess the effect of choice sequence on punishment choices. Here, choice sequence did have an effect. Subjects who made reward choices first were more willing to accept immediate punishments than subjects who had no prior experience ($F = 5.28, df = 1, 46, p < .05$ for Conditions I and II; $F = 4.75, df = 1, 46, p < .05$ for Conditions III and IV). The interactions between choice sequence and delay time and choice sequence and probability were not significant ($F_s = .22, 1.63$). The main effects of delay time and probability already have been discussed.

Relationship between Choices of Rewards and Punishments

To assess the relationship between willingness to wait for larger delayed rewards and willingness to accept smaller immediate punishments six correlation coefficients were computed. These were between reward and punishment choices when probability was 1.0 and delay times were 1, 7, and 30 days, and between the same choices when delay time was 7 days and probabilities were 1.0, .5, and .1. Only one of these six correlation coefficients was significant. When probability was .1 and delay time was 7 days, a correlation of .31 ($p < .05$) was obtained between preference for delayed larger rewards and immediate smaller punishments. The other coefficients ranged from $-.02$ to $.26$ and were not significant.

DISCUSSION

The results showed the predicted effects of the probability of the occurrence of delayed outcomes on the frequency with which they are chosen. Delayed larger rewards were chosen more frequently when the probability of obtaining them was increased. Moreover,

preference for smaller immediate punishments increased as the probability for the occurrence of the larger delayed punishment became greater. These data provide a direct confirmation of the hypothesized effects of probability on delay behavior. The results are consistent with the findings obtained by Mahrer (1956) and Mischel and Staub (1965) who studied delay of rewards but not of punishments. These investigators varied prechoice experiences designed to change subjective probability, but did not manipulate probability directly.

The present data also showed that the length of delay interval affected preference for delayed rewards in the predicted direction, with less willingness to delay for positive rewards as waiting time increased, and support earlier findings (Mischel & Metzner, 1962). However, contrary to prediction, the length of delay interval did not affect choices between immediate and delayed punishments. Children were as willing to accept immediate punishments when the waiting interval for the delayed negative outcome was small as when it was large.

Examination of mean immediate punishment choices (Table 2) indicates that subjects were close to the maximum in their willingness to accept immediate punishment at each of the three delay intervals. A possible ceiling effect on choices may thus have mitigated against obtaining the predicted effect. To test this possibility, subjects whose mean immediate negative choices across the three delay intervals exceeded 3.0 were eliminated for a subanalysis. These means for the 19 remaining subjects were 2.37, 2.16, and 1.95 at the short, intermediate, and long delay interval, respectively. Analysis of variance on these data revealed no effect of delay time even approaching significance ($F = .88$). In contrast, a comparable analysis for the effects of delay time on reward choices, using this same sample of subjects (i.e., those whose mean immediate punishment choices did not exceed 3.0), yielded a highly significant effect ($F = 8.49, p < .001$). It thus can be concluded that the amount of delay time did not affect preference for immediate negative outcomes whereas it did change delay of reward behavior.

These findings raise serious questions about the role of time in voluntary delay behavior. It has been assumed previously (Mahrer, 1956; Mischel & Metzner, 1962) that increasing delay time produces decrements in delayed reward choices by decreasing subjective probability for attainment of the delayed reward. To the extent that increasing delay time does decrease subjective probability one would expect the length of the delay interval to affect choices for both rewards and punishments. This seems particularly plausible in light of the present evidence that experimentally presented probability was a potent determinant of delay behavior for both positive and negative outcomes. The fact that the amount of delay time did not affect willingness to accept immediate punishments suggests that the duration of the delay interval either may not influence subjective probability for punishments while it does influence expectancies for rewards, or that its effects may be independent of subjective probability for both rewards and punishments. It is also possible that the effect of delay time on choice behavior may be mediated by changes produced in the subjective value of outcomes as a function of their associated delay time. This can only be answered by research on the effects of delay time on the subjective aversiveness of punishments and the subjective value of rewards.

The present results seem consistent with data from quite different contexts indicating that the temporal delay of punishment may not affect subjects' reactions to the punishment. Thus, Setterington and Walters (1964) found that problem solving in children was adversely affected by increasing delay of reward intervals (0, 10, and 20 seconds), but not by similar increases in delay of punishment intervals. In a very recent study, Elliott (1966), using short time intervals, showed that heart rate and ratings of tension by adults awaiting shock were unaffected by knowledge of the time at which the shock would arrive, consistent with earlier findings by Deane (1961). Finally, the fact that Cook and Barnes (1964), again using short delay intervals (0-8 seconds), found that adult subjects tended to choose shorter, rather than slightly longer, delay of inevitable shock

supports the view that punishments do not become less aversive as the delay time associated with them increases.

In the present study, it is of some interest that subjects who chose in a sequence of decreasing delay times were more willing to wait for delayed rewards than those who were exposed to the sequence of increasing delay times. This finding suggests that the effects of temporal intervals on reward choices may be mediated by the sequence in which these intervals occur. It is possible that the delay times with which subjects are initially confronted in a choice sequence serve as anchors or referents against which the subsequent delay intervals are contrasted and evaluated. Subjects who began with relatively long delay times may have found the shorter delay times which followed relatively less noxious than those who were exposed to increasingly long delay periods, and hence chose more delayed rewards.

The correlations obtained between choices of delayed rewards and of immediate punishments were small and mostly negligible, reaching significance in only one of six comparisons. These data suggest that willingness to wait for larger delayed rewards and to accept smaller immediate punishments may be largely unrelated and controlled by different antecedents and determinants. This lack of relationship is also consistent with the previously discussed differences in the way time appears to affect choices for delayed rewards and punishments, to the degree that it suggests different determinants for these two types of delay behavior. Since the present data constitute the only currently available evidence on the relationship between voluntary delay of rewards and punishments, it would be premature to draw strong conclusions. Nevertheless, the low interrelationships obtained argue against the notion of a unitary trait of "ability to delay" for both positive and negative outcomes.

The data on the effects of choice sequence seem especially interesting. Choice sequence did not affect preference for delayed rewards. However, subjects who had the presumably positive experience of choosing between rewards before making their choices between punishments were more willing to accept

smaller punishments immediately than those who had no prior experience. That is, subjects who had no prior experience more frequently deterred immediate punishments and chose delayed punishments of greater magnitude, regardless of the delay time or probability associated with the delayed outcome.² These data suggest that the individual's subjective state influences his subsequent choices between punishments, with greater acceptance of immediate punishment following a positive or rewarding experience. On the other hand, there was no evidence that an initial negative experience (punishment choices) facilitates the ability to delay rewards. The findings seem to support the view that rewarding experiences strengthen the organism's tolerance for undergoing immediate aversive outcomes.

² The investigation of the effects of prior rewarding and punishing experiences on choices was limited in the present design by the absence of groups in which reward experiences preceded reward choices, and in which negative or punishment experiences preceded punishment choices.

REFERENCES

- BANDURA, A., & MISCHEL, W. Modification of self-imposed delay of reward through exposure to live and symbolic models. *Journal of Personality and Social Psychology*, 1965, 2, 698-705.
- COOK, J. O., & BARNES, L. W. Choice of delay of inevitable shock. *Journal of Abnormal and Social Psychology*, 1964, 68, 669-672.
- DEANE, G. E. Human heart rate response during experimentally induced anxiety. *Journal of Experimental Psychology*, 1961, 61, 489-493.
- ELLIOTT, R. Effects of uncertainty about the nature and advent of a noxious stimulus (shock) upon heart rate. *Journal of Personality and Social Psychology*, 1966, 3, 353-356.
- FREUD, S. Formulations regarding the two principles of mental functioning. (Orig. publ. 1911) In, *Collected papers*. Vol. 4. New York: Basic Books, 1959. Pp. 13-21.
- MAHRE, A. R. The role of expectancy in delayed reinforcement. *Journal of Experimental Psychology*, 1956, 52, 101-105.
- MISCHEL, H. Trust and delay of gratification. Unpublished doctoral dissertation, Harvard University, 1963.
- MISCHEL, W. Preference for delayed reinforcement: An experimental study of a cultural observation. *Journal of Abnormal and Social Psychology*, 1958, 56, 57-61.
- MISCHEL, W. Delay of gratification, need for achievement, and acquiescence in another culture. *Journal of Abnormal and Social Psychology*, 1961, 62, 543-552. (a)
- MISCHEL, W. Father-absence and delay of gratification: Cross-cultural comparisons. *Journal of Abnormal and Social Psychology*, 1961, 63, 116-124. (b)
- MISCHEL, W. Preference for delayed reinforcement and social responsibility. *Journal of Abnormal and Social Psychology*, 1961, 62, 1-7. (c)
- MISCHEL, W. Theory and research on the antecedents of self-imposed delay of reward. In B. A. Maher (Ed.), *Progress in experimental personality research*. Vol. 3. New York: Academic Press, 1966, in press.
- MISCHEL, W., & METZNER, R. Preference for delayed reward as a function of age, intelligence, and length of delay interval. *Journal of Abnormal and Social Psychology*, 1962, 64, 425-431.
- MISCHEL, W., & STAUB, E. Effects of expectancy on waiting and working for larger rewards. *Journal of Personality and Social Psychology*, 1965, 2, 625-633.
- MOWLER, O. H. *Learning theory and behavior*. New York: Wiley, 1960.
- RAPAPORT, D. On the psychoanalytic theory of thinking. *International Journal of Psychoanalysis*, 1950, 31, 161-170.
- RENNER, K. E. Delay of reinforcement: A historical review. *Psychological Bulletin*, 1964, 61, 341-361.
- ROTTER, J. B. *Social learning and clinical psychology*. Englewood Cliffs, N. J.: Prentice-Hall, 1954.
- SETTINGER, R. G., & WALTERS, R. H. Effects of magnitude of reward and percentage of reinforcement on a lever movement response. *Child Development*, 1964, 35, 275-280.
- SINGER, J. L. Delayed gratification and ego-development: Implications for clinical and experimental research. *Journal of Consulting Psychology*, 1955, 19, 259-266.

(Received January 20, 1966)

EFFECT OF PSYCHOPATHOLOGY ON GROUP CONSENSUS AND COOPERATIVE CHOICE IN A SIX-PERSON GAME¹

V. EDWIN BIXENSTINE AND JOAN DOUGLAS

Kent State University

96 male Ss in squads of 6 made 40 simultaneous choices in a 2-choice, 6-person game of the Prisoner's Dilemma type. According to a standard personality measure, $\frac{1}{2}$ were free of psychopathology, $\frac{1}{2}$ had scores indicative of psychopathology. After 20 choices, $\frac{1}{2}$ of each group of squads above were allowed a period of free communication. The main finding was that pathology-indicative, unlike pathology-free, squads failed to make use of the communication period to establish cooperative agreements of benefit to all. They behaved no differently from squads not allowed to communicate.

A class of mixed-motive (Schelling, 1960) or non-zero-sum game known as the Prisoner's Dilemma (consult Luce & Raiffa, 1957; Scodel, Minas, Ratoosh, & Lipitz, 1959; Wilson & Bixenstine, 1962, for detailed discussions) has excited a number of researches too numerous now to be recapitulated here. For present purposes we may note that the game heretofore has been rendered for two persons (see Figure 1) and allows both to simultaneously gain or lose. With reiterated play a conflict emerges between short-term strategy and long-term strategy as is so often the case in everyday social intercourse. So far, the essential finding is that without any communication avenues other than choice sequence participants fall into a competitive, short-term strategy which restricts gain far below that potentially available.²

Recently, Bixenstine, Levitt, and Wilson (1966) have transformed this game (see Figure 2) from its customary two-person mode to a six-person (and, in principle, higher order) mode in order to explore whether a two-person game encourages more competitive behavior than does a six-person game. They also examined the influence of a free-communication period and of anonymity of choice, the

latter made possible because more than two players determine the momentary payoff values. They had squads of six subjects reiterate simultaneous choices of red (R) or blue (B) using the payoff matrix in Figure 2 under one of two conditions; namely, half of the squads had knowledge of each member's choices and half did not. These groupings were further divided into halves, one of which was allowed to communicate freely for 15 minutes while the other half was not. Thereafter, another 20 trials were administered. They found, contrary to expectation, that the level of cooperation during the first 20 trials differed not at all from that found in comparable two-person Prisoner's Dilemma games. However, the communication period resulted in dramatic mutual cooperation, but only when each subject had knowledge of each other's specific choice. The same measure of cooperation has not occurred as a function of free communication with comparable two-choice games (Bielefeld, 1964; Scodel et al., 1959).

The shift in level of cooperation following communication was pronounced and, yet, sensitive to the condition (knowledge of each other's choice) which appears to be necessary to insure trust and mutual confidence. It is often observed that psychopathology is a disturbance of the same kind, that is, a disturbance in interpersonal communication and trust. In this study, the same procedures as in the above were followed except that knowledge of choice was present in all conditions, and squads of pathology-free (PF) subjects and squads of pathology-indicative (PI) subjects

¹ This research was done under a grant (M 329-1C2) from the National Institute of Mental Health of the National Institutes of Health.

² Rapoport and Chammah (1965) report studies using a very long (300 or more) series of trials with Prisoner's Dilemma which find that after an initial decay, collaboration in some subject pairs becomes quite prominent, while in others competition continues to reign. The above generalization, however, holds for a series of trials less than about 100.

		Person II	
		Red	Blue
Person I	Red	8, 8	1, 10
	Blue	10, 1	2, 2

FIG. 1. An example of a Prisoner's Dilemma game. (The first entry in each cell is the return, usually in cents, to Person I; the second entry represents gain in cents to Person II. Players choose simultaneously and may not communicate during reiterated play of the game.)

were selected. Thus, we asked whether there were measurable differences in choice between these homogeneous squads both before and after a free-communication period.

METHOD

Subjects

Ninety-six male students enrolled at Kent State University served as subjects. They were drawn from an original volunteer group of 253 males who were administered the MMPI, short form. Thirty persons were eliminated because of questionable validity of their tests. (No subject had a validity scale *T* score above 69.) Thirty-seven were eliminated because they had only one of the pathology scales (Scales 1 through 9 excluding 5) above a *T* score of 69. Of the remainder, 124 persons had no scales above 69 and were classified as PF. From these, 48 persons were drawn at random to serve as subjects. Sixty-two persons had two or more scales (other than Scale 5) above 69 and were classified as PI. Forty-eight of these, drawn at random, served as subjects.

Apparatus

The experimental room was $10\frac{1}{2} \times 18$ feet and well illuminated. It contained a $3\frac{1}{2} \times 8\frac{1}{2}$ foot con-

Ratio of Group Choice	6A	5B/1B	4B/2B	3B/3B	2B/4B	1B/5B	6B
Choice	R B	R B	R B	R B	R B	R B	R B
Pay-off in Cents	7 a	5 11 b c	4 7 d e	3 5 f g	2 3 h i	1 2 j k	1 l
Total Gain for Group	42	36	30	24	16	11	6

FIG. 2. Payoff matrix for a six-person Prisoner's Dilemma game employed in this study. (A general statement of this game matrix observes the following relationships between payoffs: $c > e > g > i > k > l$, $a > b > d > f > h > j$, $c + e + g + i + k > a + b + f + h + j$, and $6a > 5b + c > 4d + 2e > 3f + 3g > 2h + 4i > 1j + 5k > 6l$.)

ference table upon which at equal intervals were placed six wooden control boxes, 1 foot on each side. The front and top of the boxes were cut away, and recessed within was a $\frac{1}{2}$ -inch Masonite panel, in the center of which was a $\frac{3}{8}$ -inch Weston 301 DC (0-25) milliammeter on which the 'squad's' choices were indicated. On top of each subject's box was a red and blue light which could indicate his choice. On the box was a switch which enabled the subject to register one of two choices, R or B, and affixed near the switch was a copy of the payoff matrix (see Figure 2). A response sheet was also provided which allowed subjects to record their choices.

The milliammeters were wired in series. When a panel switch was thrown left (red), all meters registered a current from a 6-volt DC source which flowed through a 144 ohm resistance. The six resistances were wired in parallel. When all six switches were on red, 25 milliamperes of current flowed, and the meters registered a full deflection. Over the face of each meter was placed a cardboard cutout with seven equally spaced ratios from 0 red/6 blue (at 0 milliamperes) to 6 red/0 blue (at 25 milliamperes) so that the meter was read by the viewers in terms of group choice ratios rather than milliamperes. The experimenter had a master control which made it possible to turn on and off the group meters and the individual lights independently. Also present was a tape recorder which was utilized during the intermission.

Procedure

After giving the MMPI short forms (which eliminates those questions located at the end of the long form not contributing to the nine scale score), the experimenter told participants that they would be called upon to take part in a later study depending upon time availability and that they would be remunerated for participating. Thereafter, the MMPI answer sheets were scored according to the manual instructions (Hathaway & McKinley, 1951).

PF and PI classifications were formed, and eight squads of six PF and eight squads of six PI subjects, drawn at random, were asked to report at the experimental room at specified times. No subject had prior knowledge of the nature of the experiment (according to response on inquiry), and none was intimately acquainted with any other subject prior to the experiment. Members of a squad were seated randomly around the table each at one of the six paneled boxes. Instructions were given which told that this was a decision-making experiment. Words such as "playing," "game," "winning," etc., were avoided. The payoff matrix (Figure 2) was explained, and the use of the tally sheet was illustrated. Subjects were asked to mark their choices (R or B) at a signal from the experimenter, then execute their choices, whereupon the experimenter would switch on all meters and individual choice lights, and finally each subject tallied his payoff in order to keep a constant visual check on his earnings. A few (three to six) practice trials were administered to familiarize subjects with the operations

of the meters, switches, and lights. A final check was made of each subject to insure that all understood clearly the nature of the payoff matrix and trial procedures. Finally, subjects were reminded that the money earned belonged to them and would be distributed at the conclusion of the experiment.

All subjects were then given 20 trials, one about every 45 seconds. At the end of the period, half of each class, PF and PI, were given a 10-minute break³ during which they were free to communicate as they wished. The experimenter left the room after turning on the tape recorder, previously adjusted and made ready. The other half of each class, PF and PI, were given a 10-minute break under instructions not to talk. During this time, a musical selection was played on the tape playback (from *The Old Man and the Sea*) while the experimenter remained in the room. Following the break, all squads returned to an additional 20 reiterations of the game as before.

At the end of the game, subjects were asked to fill out a questionnaire regarding their judgments of the game, their strategies, etc. Additionally, Firo-B (Schutz, 1958) was administered. This test was designed to measure three factors bearing on interpersonal relations (inclusion, control, and affection) and was included because of potential relationships with subjects' behaviors. Finally, subjects were paid and asked not to discuss the experiment with others as they might be potential subjects.

RESULTS

Choice Behavior

The choices exercised by the groups are represented graphically in Figure 3. The level of cooperative play (% of R chosen) is approximately uniform from the first 20 to the second 20 trial blocks for all conditions except PF communication where it makes a radical climb in the second block (following the communication period).

Since analysis on similar data indicated nonindependence of squad members, squad means were used as the base. Percentage of R chosen by 20 trial blocks for each of the 16 squads was transformed to arc sign to insure homogeneity of variance and was then submitted to an analysis of variance. Table 1 contains the summary of this analysis. Inspection of this table indicates that the interaction visible in Figure 3 between pathology

³ This period was originally designed to be 10 minutes. However, due to a misunderstanding, there was a range of from 7 to 10 minutes in the duration of the break for both PF and PI squads. As the average times and variances were essentially the same across PF and PI conditions (as will be evident in data presented later), this departure was judged to be unimportant.

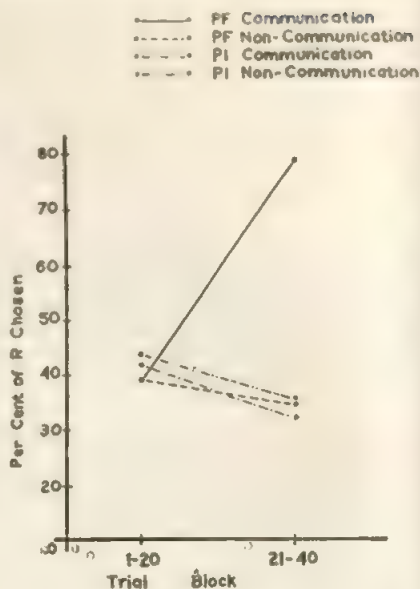


FIG. 3. Percentage of R chosen (cooperation) for the four experimental groups across trial blocks.

and communication is quite significant. It would appear that the PI groups are unable to profit from the communication period behaving as if it had not occurred. Only PF ("normals") squads made use of the period to establish mutually profitable collaboration.

Communication

Table 2 presents a breakdown of the communication period as recorded on magnetic tape. Note that most squads did not use all of

TABLE 1
ANALYSIS OF VARIANCE SUMMARY

Source	df	MS	F
Between squads	15		
Communication (A)	1	.4729	3.163*
Pathology (B)	1	.4876	3.262**
A × B	1	.7472	4.998***
Error	12	.1495	
Within squads	16		
Block (C)	1	.1288	1.573
A × C	1	.5644	6.691****
B × C	1	.7533	9.918*****
A × B × C	1	.5969	7.288*****
Error	12	.0819	

* $p < .20$.

** $p < .10$.

*** $p < .05$.

**** $p < .02$.

***** $p < .01$.

TABLE 2

DIVISION OF TIME DURING THE COMMUNICATION PERIOD BY TOPIC AND GROUP

	Time allotted (in min.)		Time in conversation (in min.)	Proportion of conversation on game
PF	1	7	6	1.00
Communication	2	10	8	1.00
Squads	3	9	6	.83
	4	9	9	.94
PI	1	8.5	8	.56
Communication	2	7	5	.75
Squads	3	10	10	.40
	4	10	10	.70

their time in conversation nor all of their conversation time about the game. The Mann-Whitney U-test (Siegel, 1956) indicates that the PF squads spent significantly more time talking about the game, relative to other topics, than did the PI squads. From listening to the tapes, it is clear that all squads were cognizant of the need for a group plan or agreement in order to maximally profit from the game. In every squad at least one person suggested a strategy to this end. However, in the PI squads, this never developed into a group agreement, while in three of the four PF squads it did.

Other Observations

A tally of the postgame questionnaire revealed that all subjects in the PF group (except the squad not reaching an agreement) indicated that the communication period greatly influenced their choice. They also indicated satisfaction with their squad and with themselves. Subjects in the PI group felt the communication period had little influence. They and subjects in the noncommunication group were variable in their expression of satisfaction in their squad and in themselves.

Seven questionnaire answers, the six scores from Firo-B, 12 scales from the MMPI (Scales 1 through 9, and *L*, *F*, and *K*), age of the subject, and the amount of R choice in the first and in the second blocks were intercorrelated (using the Pearson product-moment formula except for two questionnaire items which were dichotomous and called for biserial correlations). Correlations were computed for all 96 subjects and for the PF and

PI groups (48 subjects in each) separately. The lack of independence of subjects as regards choice behavior casts some doubt on this procedure. On the other hand, such non-independence in all likelihood worked to reduce relationships between choice behavior and personality measures; that is, insofar as squad factors determined choice, personality and other extrasquad factors are diminished in influence. Consequently, it is unlikely that this procedure would produce spuriously high relationships.

Of considerable interest are the relationships which emerged between R choice in the first 20 trial blocks and extragame measures. Table 3 carries these correlations. For all subjects, early efforts at collaboration were associated with a desire to maximize personal gain rather than to maximize difference in gain relative to the others' gain, with a tendency to feel angry with others at some time during the trials, with an elevation in the Schizophrenia scale of the MMPI, with a likelihood of choosing cooperatively in the second half (regardless of whether a communication period intervened), with a lesser desire for affection from others, with a pursuit of a cooperative strategy, and with dissatisfaction with the squad's choices. Additionally, we may note that PF subjects tend to be elevated on the Feminine Interests scale (MMPI) and to express dissatisfaction with their own choice behavior. PI subjects, in contrast, have a higher elevation on the Psychasthenia scale. It would appear that the person who tends to choose cooperatively is rather critical, somewhat distant or impersonal, and, perhaps, on the sensitive side.

Cooperative choice in the second half was significantly related to the subject's judgment that both he and his squad were pursuing a cooperative strategy, to a tendency to be angry at some time during the trials, and to R choice in the first block—nothing very surprising in view of relationships already noted. The only exceptional correlation was that with age ($r = .233, p < .05$); older subjects tended to cooperate more in the second half. Subject's age was found to have a low correlation with an acknowledgement of a cooperative strategy ($r = .215, p < .05$), an expression of a need to include others ($r =$

TABLE 3

CORRELATIONS BETWEEN R CHOICES FOR THE FIRST 20 TRIALS AND OTHER MEASURES
FOR PF, PI, AND BOTH GROUPS JOINED

	Maximize gain over difference	Anger at others dur- ing game	MMPI			R choice for 2nd block	Firo-B Wish for Affection	Dissatis- faction with self about game	Coopera- tiveness of strategy used	Dissatis- faction with squad about game
			<i>Mr</i>	<i>Pt</i>	<i>Sc</i>					
PF	.304*	.198	.321*	-.016	.200	.318*	-.091	.287*	.307*	.164
PI	.250	.412*	.084	.292*	.244	.494*	-.296*	.010	.343*	.232
Both	.252*	.321**	.186	.179	.219*	.331**	-.209*	.146	.285**	.217

* $p < .05$.

** $p < .01$.

.233, $p < .05$), and *K* on the MMPI ($r = .202$, $p < .05$). It was negatively correlated with *D* ($-.387$), *Hy* ($-.217$), *Pd* ($-.269$), *Pt* ($-.359$), and *Sc* ($-.303$) of the MMPI. There was the suggestion in this that age was related to sociability and to personal stability both of which contribute to the formation of cooperative contracts. In view of the correlations with the several MMPI scales, we were not surprised to find that our PF group was 3.2 years older (mean of 25.75 years) than the PI group ($t = 3.894$, $df = 94$, $p < .01$). This had escaped our notice previously as inspection had indicated essential age equivalence in mean and range (PF range from 18 to 46 years, PI from 18 to 43 years). Apparently, however, in selecting our MMPI PF persons, we also filtered out younger for somewhat older individuals. It seems unlikely to us that an absolute control for this slight if significant age differential would have altered our results. Still, this poses the question whether factors associated with age and not measured by the MMPI, such as wisdom or experience, might have played a part.

Of interest was the fact that rated dissatisfaction with the squad's behavior, which was correlated with R choice in the first half, was also correlated with elevated *D*, *Pd*, *Pt*, and *Sc* on the MMPI scales. Additionally, it was correlated almost significantly (.242) for the PF group with the acknowledged competitiveness of strategy, but bore a significant negative r ($-.346$) for the PI group with this rating. This same kind of difference was seen in the relationship of dissatisfaction with self and acknowledged competitiveness of strategy: PF subjects correlated positively (.279, $p < .05$), while PI lacked any correlation

($-.002$). This suggests that PF ("normal") subjects related their own competitiveness to competitiveness on the part of their squad with consequent dissatisfaction both with self and the squad, but PI subjects deny such a relationship. Further evidence for this is to be found in a significant r (.354) for PF subjects between self-rated competitiveness of strategy and assessed competitiveness of strategy of the squad, which relationship did not materialize for PI subjects ($r = -.086$).

DISCUSSION

The attempt to relate personality variables to game behavior has been marked by recurrent disappointment. Relationships established at one time or by one investigator have often failed to reappear at another time or by another investigator. However, we may have been looking into the wrong game functions. Responses to the Prisoner's Dilemma game are remarkably uniform and game specific. Particular characteristics of the player have limited opportunity to be revealed in choice behavior (see Bixenstine & Blundell, 1966). On the other hand, the ways the subject may make use of a communication period are all but unlimited. It is as if reiteration of the game as such tells us essentially about the game and its distinctive characteristics. But introducing communication transforms the game into a measure of the character of the players. Certainly, the inability of the pathology squads to establish a collaborative agreement is stark in comparison with other personality-determined game behaviors.

Just why do the PI squads fail? Based upon the recorded conversations, we ascribe no credence to the possibility that they mis-

understood or failed to grasp what was called for. In each of these squads, someone pointed out the benefits of a cooperative plan, yet this was not carried out. One member of a PI squad voiced clearly a mistrustful and rejecting reaction to a collaborative proposal indicating that each person would probably take advantage of the other's trust were such a contract established. What is more noteworthy is that little in the way of persuasion or reassurance was forthcoming from the PI members to offset this kind of mistrust, and mistrust appeared as an unavoidable reality. This interpretation, incidentally, is supported by the finding that PI subjects failed to observe a relationship between their own choice behavior and their satisfaction with the group. PF ("normals") conceivably are more optimistic about their capacity to trust the other. They are, concurrently, more likely to relate the group's failure to collaborate to their own (competitive) behavior rather than be resigned to the frustrations of mutual distrust. If they ascribe the other's mistrust to their own acts, by the same token, they would tend to assume they had control over the other's trust and collaboration.

The correlational analysis which we performed was useful in suggesting hypotheses rather than resolving them. Whether those so inclined to collaborate are rather critical, sensitive, and somewhat self-sufficient or reserved in nature must await further test.

We believe this experiment demonstrated the usefulness of groups larger than two in the study of game behavior. The Prisoner's Dilemma with six persons appeared to be more clearly problematic and an even more emphatic challenge for collaborative resolution than was the case with two players. In

closing, we would note that this study invites a further investigation of the nature of communication, contract formation, and the role of psychopathology therein. While in general the study lends support to the assertion that psychopathology is, in essence, a disruption of the communication process in interpersonal relations, we believe more precise statements regarding this disturbance are both possible and necessary.

REFERENCES

- BIELEFELD, M. The effect of communication and psychopathology on cooperation in a two-person game. Unpublished master's thesis, Kent State University, 1964.
- BIXENSTINE, V. E., & BLUNDELL, H. Control of choice exerted by structural factors in two-person, non-zero sum games. *Journal of Conflict Resolution*, 1966, 10, 478-487.
- BIXENSTINE, V. E., BEVITT, C. A., & WILSON, K. V. Decision as a function of interpersonal communication and knowledge of choice in a six-person Prisoner's Dilemma game. *Journal of Conflict Resolution*, 1966, 10, 488-496.
- HATHAWAY, S. R., & MCKINLEY, J. C. *Manual for the Minnesota Multiphasic Personality Inventory*. New York: Psychological Corporation, 1951.
- LUCE, R. D., & RAIFFA, H. *Games and decisions*. New York: Wiley, 1957.
- RAPOPORT, A., & CHAMMAH, A. *Prisoner's dilemma*. Ann Arbor: University of Michigan Press, 1965.
- SCHELLING, T. C. *The strategy of conflict*. Cambridge: Harvard University Press, 1960.
- SCHUTZ, W. C. *A three-dimensional theory of interpersonal behavior*. New York: Rinehart, 1958.
- SCODEL, A., MINAS, J. S., RATOOSH, P., & LIPETZ, M. Some descriptive aspects of two-person, non-zero sum games. *Journal of Conflict Resolution*, 1959, 3, 114-119.
- SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.
- WILSON, K. V., & BIXENSTINE, V. E. Forms of social control in two-person, two-choice games. *Behavioral Science*, 1962, 7, 92-102.

(Received February 24, 1966)

A COGNITIVE DETERMINANT OF IMPRESSION FORMATION¹

ROBERT WARE AND O. J. HARVEY

University of Colorado

Following each of 3 levels of positive and negative information about a target person, concrete and abstract Ss were assessed on the extent to which the induced impressions were generalized to other positive and negative attributes of the person and on the certainty of the generalizations. When the inputs and the generalized attributes were hedonically consistent (i.e., both were positive or both were negative), concrete Ss generalized the induced impressions further than the abstract Ss. When the input and generalized attributes were hedonically inconsistent (i.e., 1 was positive, while the other was negative or vice versa), the reverse was true. Under both conditions, however, concrete Ss were more certain of their impressions than were abstract Ss. Thus, while concrete Ss made up their minds about another person more quickly and with greater certainty than abstract Ss, they simultaneously manifested greater need for cognitive consistency.

While several studies have considered the role of personality factors in the accuracy of person perception and the attribution of certain characteristics to others (see Shrauger & Altracchi, 1964, for a review of such studies), seemingly few, if any, have been concerned specifically with the influence of personality on the amount of information sought about another person before forming an impression of him or on the certainty with which an inference is made. It is with these related questions that the present study deals.

The problem can be approached through either of two general designs. Individuals differing in personality can be left free with respect to the amount of information they seek before drawing an inference of a specified level of generality and certainty. Or the same sample of subjects can be exposed to similar amounts of information about an object person, and differences in the extent and certainty of the impression produced by common informational inputs can be determined. Because of its more ready adaptation to group administration, the latter approach was followed in this experiment.

Of the many predispositional factors that should affect certainty and extent of impres-

sion generated by controlled amounts of information, the degree of concreteness-abstractness of the conceptual system through which inputs are mediated and transmuted into psychological significance should be among the more important. The more abstract and cognitively complex individual, with his greater tendency toward processing and construing the world multidimensionally, should seek more information than the more concrete and cognitively simple person before forming an impression of high certainty about an object person. Applied to the method of the present study, this suggests that the more abstract person, from controlled amounts of information, should generalize less and hold more tentative and less certain impressions of the object person than the more concrete individual. A test of this general hypothesis was the primary concern of the study; it had the dual objective of simultaneously contributing to the accumulating information on the role of personality factors in person perception and testing some of the basic propositions underlying the theory of Harvey, Hunt, and Schroder (1961).

Concreteness-abstractness refers to a general, and presumably more or less standardized, way an individual articulates and organizes relevant aspects of his environment (Harvey et al., 1961). From a series of studies we have found greater concreteness of these mediational processes, in contrast to

¹This study was supported financially by the Group Psychology Branch, Office of Naval Research, Contract Nonr 1147 (07), with the University of Colorado. Reproduction of this study in whole or part is permitted for any purposes of the United States Government.

greater abstractness, to be shown in several ways, including (a) a simpler cognitive structure, comprised of fewer differentiations and more incomplete integrations of certain concept domains (Harvey, 1966; Harvey, Reich, & Wyer²; Harvey, Wyer, & Hautoluoma, 1963), (b) a greater tendency toward polarized evaluations, namely, good-bad, right-wrong, etc. (White & Harvey, 1965), (c) a greater dependence on authority-related cues as guidelines to belief and action (Harvey, 1964; Tiemann, 1965), (d) a greater intolerance of ambiguity, expressed in higher scores on such measures as the F and D Scales and in the tendency to form judgments of a novel situation more quickly (Harvey, 1965), (e) a greater need or tendency toward cognitive consistency and greater arousal and change from the experience of cognitive dissonance (Harvey, 1965), (f) a greater inability to change set and hence greater stereotypy in the solution of more complex and changing problems (Felkner & Harvey, 1963; Harvey, 1965), (g) a poorer delineation between means and ends and hence a paucity of different methods of solving a problem or achieving a goal (Harvey, 1965), (h) a poorer capacity to "act as if," to assume the role of the other or to think and act in terms of a hypothetical situation (Harvey, 1963; Harvey & Kline, 1965), and (i) holding of opinions with greater strength and expression of greater certainty that the opinions will not change with time (Hoffmeister³). Greater abstractness implies the reverse quantities on the above dimensions.

The above characterizations suggest clearly that, if left free to seek as much information as they wish before forming an impression of another person at some criterion level of generalization and certainty, more concrete individuals will seek less information than the more abstract persons. This should mean that when, instead of providing each person with the information he seeks, all subjects are ex-

posed to equivalent but less than maximal amounts of information, as in the present study, more concrete individuals will form more generalized and more certain impressions of the object person than will more abstract individuals. The validity of this latter assumption, however, is clearly contingent upon the level of information at which impressions are assessed.

Results from the numerous studies of the effects of intraorganismic variables upon cognition indicate that the influence of concreteness-abstractness should be greater under the ambiguity resulting from little information than under the stimulus compulsion embodied in very large amounts of information. The validity of this assumption, at least as far as this study is concerned, in turn rests upon the method by which the treatment effects are assessed. Given the condition of *hedonic consistency* between characteristics of the informational inputs and attributes on which impressions are assessed (i.e., when inputs and impression attributes are either both positive or both negative), differences in concreteness-abstractness on impression formation should be reflected more clearly at low than very high levels of informational input. Under the conditions of *hedonic inconsistency* between inputs and impression characteristics (i.e., when input characteristics are positive and the response categories are negative or vice versa), however, differences between the impressions of more concrete and more abstract individuals should be greater at the high than at the low level of information.

Postulation of these method effects stems from differences between concrete and abstract individuals previously noted. The greater need of the more concrete individuals for structure, consistency, and avoidance of dissonance should result in a greater tendency on their part than on the part of the more abstract individuals to attribute only positive characteristics to "good" people and only negative characteristics to "bad" people. Applied here, this means that more concrete individuals should generalize more than more abstract persons when the input-response characteristics are of the same hedonic quality but that the reverse picture should obtain

² O. J. Harvey, J. W. Reich, and R. S. Wyer, Jr., "Some Cognitive and Affective Determinants of Differentiation," unpublished manuscript, 1966.

³ J. K. Hoffmeister, "Some Personality Determinants of Certainty of Beliefs and Probability of Change," unpublished manuscript, University of Colorado, 1965.

when the input-impression attributes are of opposite affective qualities.

Because of the preceding considerations and the desire to maximize representativeness of the informational inputs, subjects varying in concreteness-abstractness were exposed, in three steps, to two, four, and six pieces each of positive and negative information about a hypothetical person and tested after each step on their impression of the person and the certainty of their impressions. Impressions were assessed through a kind of generalization measure based on subjects' ratings of the plausibility or likelihood that, given the positive or negative input characteristics of him, the object person would possess or manifest certain other, both positive and negative, attributes as well. A certainty rating was made of each plausibility response.

Hypotheses

Under the testing condition of hedonic consistency between input and response characteristics:

1. More concrete subjects will (a) generalize more, and (b) be more certain of their impressions than the more abstract subjects, at least under the condition of least information.

Similar differences at the other two, more compelling levels of information would not negate our hypotheses, but would simply indicate that the influence of concreteness-abstractness on impression formation is stronger than we had any bases for hypothesizing.

Under the testing condition of hedonic inconsistency between input and response attributes:

2. More concrete subjects will (a) generalize less, but (b) be more certain of their impressions than the more abstract subjects, at least under the condition of greatest information.

Similar differences at the lower levels of information would simply indicate a stronger than anticipated effect of concreteness-abstractness.

No further hypotheses were made about the main or interactional effects of quality and frequency of inputs, although effects in

addition to those hypothesized were statistically analyzed and will be reported.

METHOD

Subjects

Eighteen concrete and 18 abstract subjects participated in the experiment. As students in introductory psychology at the University of Colorado during the fall of 1964, they received either course credit or cash payment for serving as subjects.

Measurement of Concreteness-Abstractness

This was accomplished by the This I Believe Test, an instrument that requires the respondent to complete in two or three sentences the phrase, "This I believe about . . .," the blank being filled successively by concept referents such as "the American way of life," "religion," "marriage," "friendship," "people," and "guilt" (Harvey, 1964, 1965; White & Harvey, 1965). From the normativeness, absolutism, evaluativeness, and simplicity-complexity of the completions, together with criteria suggested in the characterizations noted earlier, respondents may be classified into one of the four principal systems and levels of abstractness posited by Harvey et al. (1961) or into some admixture of two or more systems. Since the concern of this study was with the effects of concreteness-abstractness, without regard to content differences of the different conceptual systems, only subjects who were classified as representing predominantly System 1, the more concrete mode of conceptual functioning, and System 4, the more abstract way of functioning, were included in the experimental sample.

Stimulus Materials

Input stimuli. Impressions of the object person were elicited by brief descriptions of 12 behavioral acts to which all subjects were exposed. As determined by pretesting of a larger number of behavioral descriptions, six of the statements connoted socially desirable behavior (*positive inputs*), while the other six described socially undesirable behavior (*negative inputs*). The statements, in typewritten form, were enclosed in glass slides, two desirable or two undesirable acts described in each of the six slides.

Output dimensions. Behavioral output, or the extent to which impressions elicited by the input stimuli were generalized to other aspects of the object person, was assessed from responses to six descriptive statements, three of which had been shown by pretesting to connote socially desirable acts or attributes (*positive outputs*) and three of which had been shown to imply negative attributes (*negative outputs*). After the presentation of each slide, subjects responded to each of the six randomly ordered output descriptions in two ways: by indicating how plausible the attribute was, and by expressing their feeling of certainty about each plausibility response. Both plausibility and certainty re-

sponses were in the form of ratings made on graphic rating scales on which only the opposite poles were indicated for each behavioral dimension. "Extremely likely" and "Extremely unlikely" defined the poles for each plausibility rating, while certainty ratings were made between the poles of "Very certain" and "Very uncertain."

Procedure

Subjects were instructed:

This is an experiment in judging behavior on the basis of knowledge about some past behavior of an individual. We want to know how likely or plausible you think it is that a person who has shown certain behaviors in the past would also show certain other behavior or characteristics. Information pertaining to some behaviors of an individual will be projected upon the screen. You are to make judgments about the likelihood of plausibility that this individual would evidence other behavior or attributes indicated on your response sheet. Also, you are to indicate how strongly you feel that the plausibility judgments you make are in fact correct. Work as rapidly as possible without too much deliberation, giving only your first impressions.

Following instructions on the use of the scales, half of the subjects in each personality classification were exposed to negative inputs and then to positive inputs, making plausibility and certainty ratings after each slide; the other half of the subjects were presented the positive and negative inputs in the

TABLE 1
INPUT DESCRIPTIONS OF OBJECT PERSONS

Slide no.	Negative inputs	Slide no.	Positive inputs
1	Had high job absenteeism Ran over his neighbor's dog with his car	1	Frequently sent flowers and get-well wishes to hospitalized friends Worked his way through college by part-time employment as a laborer
2	Picked up tips left for the waitress by other patrons Court-martialed while serving in the U. S. Army	2	Spearheaded a fund-raising campaign for homes for unwed mothers Became a vice-president of a large company at an unusually young age
3	Arrested for failure to support his child Convicted for pushing dope to high school students	3	On several occasions voluntarily assumed the work responsibilities of an ill colleague Decorated for gallantry in action while serving in the U. S. Marine Corps

TABLE 2
OUTPUT DIMENSIONS

Positive outputs	Negative outputs
1. Have a strong and physically attractive body	1. Read friends' private letters without their permission
2. Suppress the urge to speak hastily in anger	2. Treat all religious matters as primitive and trivial
3. Help other people feel more secure even if he does not like them	3. Break promises without justification

reverse order. Table 1 presents the input statements and the sequence of their presentation, and Table 2 contains the six output dimensions.

RESULTS

The assumption that variation in concreteness-abstractness would dispose toward opposite generalization effects depending on whether the input and output dimensions were hedonically consistent or inconsistent necessitated the separate treatment of plausibility ratings of the three positive and the three negative output dimensions. Two generalization scores represented by the respective sums of the plausibility ratings (measured in millimeters) on the three positive and on the three negative output dimensions were consequently computed for each subject for each pair of inputs. For computational ease, all sums were divided by 10 and rounded to the nearest whole number. Certainty scores were computed and rounded in similar ways.

Hedonic Quality and Frequency of Inputs

The anticipated effects of variation in concreteness-abstractness rest upon the assumption that manipulation of the stimulus variables would produce certain effects independently of personality: namely, the positive inputs would generate impressions that were more positive than negative; negative inputs would elicit impressions more negative than positive; the tendency of inputs to produce images consistent with their hedonic quality would increase with increased amounts of information.

That these necessary stimulus effects were achieved can be inferred from Table 3.

Subjects clearly considered it more plausible that a person described in positive terms

TABLE 3

MEAN PLAUSIBILITY RATINGS OF POSITIVE AND NEGATIVE OUTPUT DIMENSIONS FOR EACH PAIR OF POSITIVE AND NEGATIVE INPUTS

No. of inputs	Positive inputs		Negative inputs	
	Positive outputs	Negative outputs	Negative outputs	Positive outputs
2	27.64	9.64	25.25	15.20
4	28.84	9.22	30.05	13.39
6	31.00	5.89	34.25	10.11

should display positive rather than negative attributes and that a negatively depicted person should behave more undesirably than positively characterized individuals. As would have to be the case if the stimulus manipulations were effective as desired, these differences were accentuated with increased frequency of informational inputs.

While the preceding results simply show effectiveness of the stimulus variations, other comparisons of the effects of the quality (i.e., positive or negative) of the inputs under the different conditions of input-output consistency and inconsistency generate more interesting results. No significant difference was found by *t* tests between the positive generalization (i.e., the plausibility ratings of positive output dimensions) produced by positive inputs with the negative generalization (i.e., the plausibility ratings of the negative output attributes) produced by the negative inputs, for the total generalization or the generalization at each input frequency. This means that the tendency to generalize from positive information about a person to other positive characteristics of him is comparable in strength to the tendency to attribute negative characteristics to another following a negative description of him.

The picture is different, however, when comparisons are made between extent of generalization under the input-output inconsistency conditions. Subjects were less willing at all levels of information to attribute negative characteristics to a person following positive depiction of him than they were to attribute positive attributes to him following a negative description, implying a kind of tendency to see the good in another person rather than the bad. This difference, of similar di-

rection for both the concrete and abstract subjects, was significant for total generalization ($t = 5.60, p < .001$) and for each of the three levels of information, the smallest *t* (for two inputs) being 5.04 ($p < .001$).

As with extent of generalization or extent of perceived plausibility, Table 4 shows that certainty of ratings became greater as the frequency of inputs increased and that there was no difference between the certainty of the ratings of the positive output attributes following positive information and the certainty of the ratings of the negative output dimensions following negative inputs ($t = .33$). Also as with generalization but in the opposite direction, there was a significant difference between the certainty of the ratings under the different conditions of input-output inconsistency. While subjects considered positive behavior following negative inputs to be more plausible than negative behavior following positive representations, they were less certain of the latter ratings; that is, they felt more certain that a person described as having positive characteristics would be less likely to manifest undesirable behavior than a negatively depicted person would be to display positive characteristics.

Concreteness-Abstractness

Given most of the preceding results, a presentation of the effects of concreteness-abstractness becomes appropriate.

Results in Table 5 indicate that when input and output characteristics are hedonically consistent, the extent of generalization or willingness to go beyond the information given in the inferences drawn about another person is greater for the concrete than for the ab-

TABLE 4

MEAN CERTAINTY OF IMPRESSION FOR EACH PAIR OF INPUTS UNDER THE CONDITIONS OF INPUT-OUTPUT CONSISTENCY AND INCONSISTENCY

No. of inputs	Positive inputs		Negative inputs	
	Positive outputs	Negative outputs	Negative outputs	Positive outputs
2	27.33	28.48	25.34	24.03
4	31.05	31.84	31.25	27.98
6	33.08	34.69	35.56	32.70

TABLE 5

MEAN GENERALIZATION OF CONCRETE AND ABSTRACT SUBJECTS FOR THREE LEVELS OF INFORMATION UNDER EACH OF THE CONDITIONS OF INPUT-OUTPUT CONSISTENCY AND INCONSISTENCY

S group	Input-output hedonically consistent			Input-output hedonically inconsistent		
	No. of inputs			No. of inputs		
	2	4	6	2	4	6
Concrete	54.56	61.94	69.22	22.06	19.89	14.06
Abstract	51.22	55.83	61.28	27.61	25.33	17.94

strat individuals. While our hypothesis stipulated that this difference should be significant at least for the lowest level of information, t tests showed the difference to be significant at all three levels of information (smallest t , for two inputs = 1.69, $p < .05$) as well as for the levels combined ($t = 2.95$, $p < .005$). Thus concreteness-abstractness exercised significant effects even under the more highly structured conditions represented by the greater frequencies of inputs.

In fact, the difference between the amounts of generalization of the concrete and abstract subjects was greater at the highest than at the lowest level of information. This is accounted for by the fact that while the concrete and abstract subjects both increased in generalization with increased inputs, as should be expected, the rate of increase, as measured by the difference in generalization between two and six inputs, was significantly greater for the concrete than for the abstract individuals ($t = 1.70$, $p < .05$). Moreover, it may be noted, as a fact of further interest, that the concrete subjects generalized as much from only four inputs as the abstract subjects did from six.

Results in Table 5 also show that under input-output inconsistency both concrete and abstract subjects decreased in generalization

with increased inputs, meaning simply that the greater the frequency of positive inputs the less the willingness or tendency to attribute negative characteristics to the object person, and the greater the frequency of negative inputs the greater the unwillingness to attribute positive features to him. This fact is of little theoretical import, but of theoretical interest is the finding that the concrete and abstract subjects differed significantly in this effect at all but the highest level of information. After each of the first and second pair of inputs (t 's = 2.33, 2.06, p 's $< .03$) as well as when the combined effects of all six inputs were considered ($t = 2.20$, $p < .03$), the concrete subjects considered it less plausible than the abstract individuals that a person who had earlier manifested either socially desirable or undesirable behavior subsequently would evidence behavior of the opposite quality.

In addition to the predicted influence of concreteness-abstractness on generalization, this variable also exercised a significant effect on certainty of the elicited impressions. Results in Table 6 show that whether the input and output attributes were consistent or inconsistent, a contingency that had no effects on degree of certainty, the concrete individuals were more certain of their impressions than were the abstract subjects (t for combined conditions = 2.04, $p < .02$).

DISCUSSION

The fact that under the condition of input-output consistency greater concreteness accompanied greater impression generalization at the highest as well as at the two lower levels of information indicates that the variable of concreteness-abstractness exercises con-

TABLE 6

MEAN CERTAINTY OF THE CONCRETE AND ABSTRACT SUBJECTS UNDER INPUT-OUTPUT CONSISTENCY AND INCONSISTENCY

S group	Input-output hedonically consistent	Input-output hedonically inconsistent
Concrete	188.78	189.44
Abstract	178.56	169.89

siderable influence on the inferences people make of others even in the presence of fairly highly structured stimulus cues. For a more thorough assessment of the effects of this cognitive factor, it is clear that it must be studied under a wider range of stimulus or cue structure than was investigated here, from less information and greater ambiguity through more frequent and more determinate informational inputs. An approach that might accomplish this and at the same time be more sensitive to personality differences would be the one proposed as an alternative to the present one, namely, leaving individuals free in the amount of information they seek before reaching an impression of a specified level of generality and/or certainty rather than exposing them to constant amounts of information and examining their differential responses on these dependent variables.

In addition to indicating that concrete individuals make up their minds more completely than abstract persons on little information, the present results also point clearly to a greater need for cognitive consistency on the part of the concrete than on the part of the abstract person. The fact that the concrete subjects generalized more than the abstract individuals under conditions of hedonic consistency between input and output dimensions, but generalized less than the abstract subjects when the inputs were hedonically inconsistent, may be interpreted as supporting the notion that concrete persons, more than abstract individuals, are disposed toward seeing the "good" person as all good and the "bad" person as all bad, minimizing the plausibility that the same person could simultaneously possess both good and bad characteristics. This interpretation would be in line with one of our earlier findings (Harvey, 1965) that concrete individuals were less

able than abstract subjects to generate superordinate constructs of persons that would be consistent with their possessing simultaneously positive and negative characteristics

REFERENCES

- FELKNOR, C., & HARVEY, O. J. Cognitive determinants of concept formation and attainment. Technical Report No. 10, 1963, University of Colorado, Contract Nonr 1147(07), Office of Naval Research.
- HARVEY, O. J. Cognitive determinants of role playing. Technical Report No. 3, 1963, University of Colorado, Contract Nonr 1147(07), Office of Naval Research.
- HARVEY, O. J. Some cognitive determinants of influenceability. *Sociometry*, 1964, 27, 208-221.
- HARVEY, O. J. Some situational and cognitive determinants of dissonance resolution. *Journal of Personality and Social Psychology*, 1965, 1, 349-355.
- HARVEY, O. J. System structure, flexibility and creativity. In O. J. Harvey (Ed.), *Experience, structure and adaptability*. New York: Springer, 1966. Pp. 39-65.
- HARVEY, O. J., HUNT, D. E., & SCHRODER, H. M. *Conceptual systems and personality organization*. New York: Wiley, 1961.
- HARVEY, O. J., & KLINE, J. A. Some situational and cognitive determinants of role playing: A replication and extension. Technical Report No. 15, 1965, University of Colorado, Contract Nonr 1147(07), Office of Naval Research.
- HARVEY, O. J., WYER, R. S., JR., & HAUTALUOMA, J. E. Some cognitive and affective determinants of differentiation and integration. Technical Report No. 8, 1963, University of Colorado, Contract Nonr 1147(07), Office of Naval Research.
- SHRAUGER, S., & ALTRACCHI, J. The personality of the perceiver as a factor in person perception. *Psychological Bulletin*, 1964, 5, 289-308.
- TIEMANN, H. A. Some social and personality determinants of reaction to sensory deprivation. Technical Report No. 14, 1965, University of Colorado, Contract Nonr 1147(07), Office of Naval Research.
- WHITE, B. J., & HARVEY, O. J. Effects of personality and own stand on judgement and production of statements about a central issue. *Journal of Experimental Social Psychology*, 1965, 1, 334-347.

(Received February 25, 1966)

PARENT DISCIPLINE AND THE CHILD'S MORAL DEVELOPMENT¹

MARTIN L. HOFFMAN

University of Michigan

AND

HERBERT D. SALTZSTEIN

Massachusetts Institute of Technology

7th-grade children were assessed on several dimensions of moral development by means of paper-and-pencil tests and ratings by parents, teachers, and peers. Extreme groups were formed along each of these dimensions, and they were compared on measures of parental discipline based on reports by the children themselves and by each of the parents. Discipline techniques were coded into 3 categories: power assertion, in which the parent capitalizes on his power and authority over the child; love withdrawal, i.e., direct but non-physical expressions of anger, disapproval, etc.; and induction, consisting of the parent's focusing on the consequences of the child's action for others. Data from middle- and lower-class boys and girls were analyzed separately. IQ was controlled for each analysis. With considerable but not complete consistency, advanced development along the various moral dimensions was associated with infrequent use of power assertion and frequent use of induction among the middle-class sample. Love withdrawal, on the other hand, related infrequently to moral development.

Recent years have seen the accumulation of a body of findings relating moral development, especially internalization of moral values and the capacity for guilt, to parental practices. In a recent review of this research (Hoffman, 1963a) the following propositions received support: (a) A moral orientation based on the fear of external detection and punishment is associated with the relatively frequent use of discipline techniques involving physical punishment and material deprivation, here called power assertive discipline; (b) a moral orientation characterized by independence of external sanctions and high guilt is associated with relatively frequent use of nonpower assertive discipline—sometimes called psychological, indirect, or love-oriented discipline.

Several explanations of these findings have been advanced, each focusing on a different aspect of the parent's discipline. Thus, Allin-smith and Greening (1955) suggest that the significant variable may be the difference in the model presented by the parent during the disciplinary encounter (i.e., parent openly expresses anger versus parent controls anger).

The importance of this factor may lie in the model it provides the child for channeling his own aggression. Where the parent himself expresses his anger openly, he thereby encourages the child to express his anger openly; where the parent controls his anger, he discourages the child from openly expressing anger and therefore may promote a turning of the anger inward which according to psychoanalytic theory is the process by which the guilt capacity is developed.

Another explanation of the difference between power assertive and nonpower assertive techniques is in terms of the duration of the punishment; that is, whereas nonpower assertive discipline may last a long time, the application of force usually dissipates the parent's anger and thus may relieve the child of his anxiety or guilt rather quickly. A third possibility, suggested by Sears, Maccoby, and Levin (1957), is that punishing the child by withholding love, which is frequently involved in nonpower assertive discipline, has the effect of intensifying the child's efforts to identify with the parent in order to assure himself of the parent's love.

A still different formulation has recently been suggested by Hill (1960). According to this view, the crucial underlying factor is the timing of the punishment. Love-withdrawal punishment is believed more often to termi-

¹This study was supported by Public Health Service Research Grant M-02333 from the National Institute of Mental Health. It was carried out while both authors were at the Merrill-Palmer Institute.

The authors wish to thank Lois W. Hoffman for her many helpful comments and suggestions.

nate when the child engages in a corrective act (e.g., confession, reparation, overt admission of guilt, etc.), whereas physical punishment is more likely to occur and terminate at the time of the deviant act and prior to any corrective act.

Finally, the important variable may be the information often communicated by nonpower assertive techniques regarding the implications of the child's deviant behavior. For example, Aronfreed's (1961) view is that such information can provide the cognitive and behavioral resources necessary for the child to examine his actions independently and accept responsibility for them.

Though varied, all but the last of these explanations assume the key ingredient for nonpower assertive discipline to be its punitive—more specifically, its love-withdrawing—quality. This hypothesis stems from psychoanalytic and learning theories that emphasize anxiety over loss of love as the necessary motivational basis for moral development.

In examining instances of nonpower assertive discipline it became apparent that the amount of love withdrawal, real or threatened, varied considerably. In some cases, the love-withdrawal aspect of the discipline seemed to predominate. In others it seemed totally absent, and in still others it seemed to be a minor part of a technique primarily focused on the harmful consequences of the child's behavior for others. This suggested that the effectiveness of these techniques might lie in their empathy-arousing capacity rather than, or in addition to, their love-withdrawing property. In the present study we accordingly made the distinction between two kinds of nonpower assertive discipline. One, called *induction*, refers to techniques in which the parent points out the painful consequences of the child's act for the parent or for others. In the second, called *love withdrawal*, the parent simply gives direct but nonphysical expression to his anger or disapproval of the child for engaging in the behavior. In a sense by these latter techniques the parent points out the painful psychological consequences of the act for the child himself, that is, the withdrawal of love by the parent.

It is probable, of course, that the child experiences both these types of nonpower as-

sertive techniques as involving a loss of love. However, as indicated above, the love-withdrawing component of the induction techniques is more subdued, and in addition they provide him with the knowledge that his actions have caused pain to others. By doing this the technique capitalizes on the child's capacity for empathy. In our view (see Hoffman, 1963b; Hoffman, in press; Hoffman & Saltzstein, 1960) it is this capacity for empathy which provides a powerful emotional and cognitive support for development of moral controls and which has been overlooked in other psychological theories of moral development. For this reason it was expected that *induction*, and not *love withdrawal*, would relate most strongly to the various indexes of moral development.

Affection has often been supposed to be a necessary condition for moral development. Measures of the parent's affection were therefore included for completeness. We expected, following the pattern of the previous research, that power assertion would relate negatively, and affection positively, to the moral indexes.

METHOD

Sample

The children studied were all seventh graders in the Detroit metropolitan area. The test battery was administered to groups of children in the schools during three sessions spaced about a week apart. Sometimes an individual class was tested in the homeroom, and sometimes several groups were tested together in the gymnasium or auditorium.

Data bearing on the various dimensions of moral development were obtained from over 800 children broadly representative of the population in the area. Because of the apprehension of some of the school officials, however, we were unable to obtain reports of parental discipline from about a fourth of these children, the loss being greater among the lower-class sample. In addition, children identified as behavior problems and those from nonintact families were screened from the sample. Further shrinkage due to absences, incomplete background information, and unintelligible or incomplete responses resulted in a final sample of 444 children. Included were 146 middle-class boys, 124 middle-class girls, 91 lower-class boys, and 83 lower-class girls.

Subsequently, interviews were conducted with a subsample consisting of 129 middle-class mothers (66 boys and 63 girls) and 75 middle-class fathers (37 boys and 38 girls). No interviews were conducted with parents of the children from the lower class.

Child Morality Indexes

Several different moral indexes were used—each tapping a different aspect of conscience.² The two major indexes pertain to the degree to which the child's moral orientation is internalized. These are (a) the intensity of guilt experienced following his own transgressions, and (b) the use of moral judgments about others which are based on internal rather than external considerations. The other indexes pertain to whether the child confesses and accepts responsibility for his misdeeds and the extent to which he shows consideration for others. Identification, though not a direct moral index, was also included because of its relationship to moral development, as hypothesized by psychoanalytic theory and by recent researchers (e.g., Sears et al., 1957).

Guilt. Two semiprojective story-completion items were used to assess the intensity of the child's guilt reaction to transgression. The technique presents the child with a story beginning which focuses on a basically sympathetic child of the same sex and age who has committed a transgression. The subject's instructions are to complete the story and tell what the protagonist thinks and feels and "what happens afterwards." The assumption made is that the child identifies with the protagonist and therefore reveals his own internal reactions (although not necessarily his overt reactions) through his completion of the story.

The first story used here was concerned with a child who through negligence contributed to the death of a younger child. The story beginning was constructed so as to provide several other characters on whom to transfer blame. The second story was about a child who cheats in a swimming race and wins. In both stories detection was made to appear unlikely. In rating the intensity of the guilt from the subject's completion of the story, care was taken to assess first that the subject identified with the central character. If such identification was dubious, the story was not coded for guilt, nor were stories involving only external detection or concern with detection coded for guilt. All other stories were coded for guilt. For a story to receive a guilt score higher than zero there had to be evidence

of a conscious self-initiated and self-critical reaction. Given evidence for such a reaction, the intensity of guilt was rated on a scale ranging from 1 to 6. At the extreme high end of the scale were stories involving personality change in the hero, suicide, etc. In coding the stories the attempt was made to ignore differences in sheer style of writing and to infer the feeling of the subject as he completed the story.

A departure from the usual practice was to assign two guilt scores to each story—one for the maximum guilt experienced by the hero, usually occurring early in the story, and the other for terminal guilt. In relating discipline to this and other facets of morality extreme groups were chosen. In choosing the high- and low-guilt groups, attention was paid to both scores. That is, the high-guilt group included those who sustained a high level of guilt throughout the stories. The low-guilt group included children who manifested little or no guilt throughout the stories. Children who initially manifested intense guilt which was dissipated through confession, reparation, defenses, etc., were not included in the guilt analysis.

Internalized moral judgments. The moral judgment items consisted of several hypothetical transgressions which the children were asked to judge. These situations were of the general type used by Piaget, including moral judgments about persons committing various crimes, for example, stealing; choosing which of two crimes was worse, for example, one involving simple theft and the other a breach of trust; and judgments of crimes with extenuating circumstances, for example, a man who steals in order to procure a drug which he cannot afford and which is needed to save his wife's life.³ In each case the child's response was coded as external (e.g., "you can get put in jail for that"), internal (e.g., "that's not right, the man trusted you"), or indeterminate. The individual internal scores were then summed for all items, and the sum constituted the child's internalization score on moral judgments.

Overt reactions to transgression. Two measures were used to assess the child's overt reactions to transgression. The first was the teacher's report of how the child typically reacts when "caught doing something wrong." The categories included: "denies he did it"; "looks for someone else to blame"; "makes excuses"; "cries, looks sad, seems to feel bad"; "accepts responsibility for what he has done"; and "where possible tries on own initiative to rectify situation."

The second measure was a questionnaire item asked of the child's mother, similar to the item used by Sears et al. (1957). The question was: "when . . . has done something that (he) (she) knows you would not approve of, and you haven't found out about it yet, how often does (he) (she) come and tell you about it without your asking?" The mother was asked to check one of five alterna-

² These dimensions were used because they clearly bear on morality and because they represent different levels (affective, cognitive, overt) and directions for behavior (proscriptions, prescriptions). Each dimension has its advantages and disadvantages, and since a strong case for including one and not the others could not be made we included them all. In doing this our intention was not to treat them as indexes of a single underlying "moral development." Doing this would seem premature, since, although the different aspects of morality presumably increase with age (empirical data on age progression are available only for moral judgment), they very likely begin to develop—and reach full development—at different ages and progress at different rates.

³ This item was an adaptation of one used by Kohlberg (1963).

tives, the extremes of which were "all the time" and "never."

Neither of these measures is ideal. The first has the disadvantage of asking for the child's reaction in the presence of an authority figure after detection. The second has the defect of being based on a report by the parent, who is the same person providing much of the discipline data and who is more likely to be influenced by "social desirability" than the teacher. Yet, the parent may well be the only person with enough background information and close contact with the child to make a knowledgeable estimate of how he acts before detection.

Consideration for other children. This measure was obtained from sociometric ratings by the children in the same classroom. Each child made three nominations for the child first, second, and third most "likely to care about the other children's feelings" and "to defend a child being made fun of by the group." The usual weights were used and the two scores summed.

Identification. Our major measure of identification was based on the child's responses to several items bearing on his orientation toward the parent: (a) admiration: "Which person do you admire or look up to the most?"; (b) desire to emulate: "Which person do you want to be like when you grow up?"; (c) perceived similarity: "Which person do you take after mostly?" Responses which mention the parent were coded as parent-identification responses and summed to obtain an overall identification score. It should be noted that this measure is designed to assess the child's conscious identification with the parents and not necessarily the unconscious identification of which Freud wrote.

Coding procedure. The story completion and moral judgment coding were done by one of the authors (HDS). To avoid contamination, the procedure was to go through all 444 records and code one item at a time. Especially difficult responses were coded independently by both authors, and discrepancies were resolved in conference.

Before the final coding was begun, coding reliabilities of 82% for maximum guilt, 73% for terminal guilt, and 91% for internal moral judgment were attained by the authors. These figures represent the percentage of agreement in giving high (top quartile), low (bottom quartile), and middle ratings. There were no extreme disagreements, that is, no instances in which a child received a high rating by one judge and a low rating by the other.

Measures of Parent Practices

Two reports of each parent's typical disciplinary practices were available—one from the children who reported the disciplinary practices of both parents, another from the mothers and fathers who each reported their own typical disciplinary practices. The reports from the children were collected during the third testing session in the schools. The parents were interviewed separately by trained female interviewers. The interview typically lasted about an hour.

Assessment of parental discipline was made in the following way. Each respondent (the child or parent) was asked to imagine four concrete situations: one in which the child delayed complying with a parental request to do something, a second in which the child was careless and destroyed something of value, a third in which he talked back to the parent, a fourth situation in which he had not done well in school. Following each situation was a list of from 10 to 14 practices. The respondent was asked to look over the list, then rate the absolute frequency of each and finally to indicate the first, second, and third practice most frequently used.⁴ These three choices were weighted, and the scores summed across the four situations. The practices listed represented our three main categories. The first category, *power assertion*, included physical punishment, deprivation of material objects or privileges, the direct application of force, or threat of any of these. The term "power assertion" is used to highlight the fact that in using these techniques the parent seeks to control the child by capitalizing on his physical power or control over material resources (Hoffman, 1960). The second category, *love withdrawal*, included techniques whereby the parent more or less openly withdraws love by ignoring the child, turning his back on the child, refusing to speak to him, explicitly stating that he dislikes the child, or isolating him. The third category, *induction regarding parents*, includes appeals to the child's guilt potential by referring to the consequences of the child's action for the parent. Included are such specifics as telling the child that his action has hurt the parent, that an object he damaged was valued by the parent, that the parent is disappointed, etc.

These lists were administered to each parent twice, once with instructions to select the techniques which he used at present, and next to select those he remembers using when the child was about 5 years old. Reports of past discipline were not asked of the children because it was unlikely that they could remember parent practices used several years before.

The above measure of induction is a limited one in that it only included instances where the parent made references to the consequences of a transgression for the parent himself. To supplement this, an additional measure of induction was constructed. This dealt with the parent's reaction to two situations in which the child's transgression had harmful consequences for another child. In the first situation the child, aged 5, aggresses against another child and destroys something the other child has built, causing the other child to cry. In the second situation the parent sees his child aged 6-10 making fun of another child. The parent was asked what he would have done or said in such a situation, and his reaction was coded along a 3-point scale for the degree to which he (the parent) makes reference to and

⁴ Ratings of the absolute frequency were included primarily to make sure the respondent thought about all the items in the list before ranking them.

shows concern for the *other* child's feelings. The scores were summed to arrive at a measure of the parent's use of *induction regarding peers*.

Assessment of the parent's affection for the child was also obtained from the child and from the parent. The child was given a list of 19 behaviors indicating affection, approval, criticism, advice giving, and participation in child-centered activities and asked to indicate along a 4-point scale how often the parent engaged in such behaviors. The affection score was a simple weighted sum for the affection and approval items.

A slightly different measure was used to obtain affection data from the parents. They were given a list of eight behaviors indicating affection, approval, qualified approval, and material reward and asked to indicate along a 4-point scale how often they engaged in such behaviors when the child "did something good." The affection score was a weighted sum for the affection items.

Background information. The family's social class was determined from the child's responses to questions about the father's occupation and education. The distinction was basically between white collar and blue collar. In a few cases, families initially classified as middle class were later recategorized as lower class as a result of more accurate and specific information from the parent about the father's actual occupation and education.

Data analysis. The data were analyzed separately for middle-class boys, middle-class girls, lower-class boys, and lower-class girls. The procedure for each of these subsamples was to form two groups—one scoring high and one scoring low on each moral

development index—and then to compare these groups on the child-rearing-practice scores obtained in the child reports and (in the case of the middle class only) the parent interviews. In forming the comparison groups, the cutoff points were made as close as possible to the upper and lower quartile points within each subsample.

The test of significance used throughout was the median test.

Control on IQ. An important feature of this study, which was not true in the previous moral development research, was the control on intellectual ability which was instituted. Scores on either the California Test of Mental Maturity or the Iowa Tests of Basic Skills were found—with social class controlled—to relate positively to internalized moral judgments and consideration for others, negatively to confession, and negatively to parent identification. This suggested that some of the findings previously reported in the literature might be the artifactual results of a lack of IQ control. In forming the high and low quartile groups for these variables we therefore controlled IQ—to the point of making the high-low differences in IQ negligible. Since IQ did not relate to guilt, there was no need to control IQ in the guilt analysis.

RESULTS AND DISCUSSION

To facilitate presentation of the results, the significant findings relating moral development indexes and parental discipline are summarized in Tables 1 and 2 for the middle-

TABLE 1
STATISTICALLY SIGNIFICANT RELATIONS BETWEEN CHILD'S MORALITY INDEXES AND
MOTHER'S DISCIPLINE TECHNIQUES: MIDDLE CLASS

Morality index	Power assertion			Love withdrawal			Induction re parent			Induction re peers ^a		
	Boys	Girls	Sum	Boys	Girls	Sum	Boys	Girls	Sum	Boys	Girls	Sum
Guilt (child's response)		-p*	-c*				+c*	+p*	+c*	+p*		+p**
			-n*						+n*			
			-p*						+p*			
			-c*		-c*	-c*		+c*				
Internal moral judgment (child's response)		-n*										
Confession (mother's report)	-p**		-p**				+n*		+c*			
Accepts responsibility (teacher's report)	-c*	-c*	-c**	+n*			+c*		+c**			
		-n*	-n**				+n*					
							+p*					
Consideration for other children (peers' ratings)	+n*	-p*		-p*				+n*	+c*		+p**	+p**
								+p*				
Identification (child's response)	-c*	-c*	-c**	-n*			+p*	+c*	+c*			

Note.—The data sources of the significant findings summarized in Tables 1, 2, and 4-6 are indicated as follows: c (child report), n (parent report of current practices), p (parent report of past practices).

^a Data on induction regarding peers are incomplete since these data were obtained only from the parent reports of past practices.

* $p < .05$.

** $p < .01$.

TABLE 2

STATISTICALLY SIGNIFICANT RELATIONS BETWEEN CHILD'S MORALITY INDEXES AND
FATHER'S DISCIPLINE TECHNIQUES: MIDDLE CLASS

Morality index	Power assertion			Love withdrawal			Induction re parent			Induction re peers		
	Boys	Girls	Sum	Boys	Girls	Sum	Boys	Girls	Sum	Boys	Girls	Sum
Guilt (child's response)												
Internal moral judgment (child's response)		-c*						+c*				
Confession (mother's report)	+p*		+p*		+c*		-p*		-p*			
Accepts responsibility (teacher's report)	-c**		-c*			+c*						
Consideration for other children (peers' ratings)	+n*		+p*		-c*			+c**	+c**			
Identification (child's response)												

* $p < .05$.

** $p < .01$.

class sample and Tables 4 and 5 for the lower-class sample.⁵ Included in each table are relationships between each of the six indexes of moral development and each of the four measures of parental discipline: power assertion, love withdrawal, induction regarding parents, and induction regarding peers. Tables 1 and 2 are based on present discipline as reported by the child and present and past discipline as reported by the parent. Since the parent's report was not available for the lower-class sample, Tables 4 and 5 are based solely on the child's report of present parental discipline.

Middle-class discipline. The overall pattern of the findings in the middle class provides considerable support for our expectations, at least with respect to the mother's practices. Thus the frequent use of power assertion by the mother is consistently associated with weak moral development. The use of induction, on the other hand, is consistently associated with advanced moral development. This is true for both induction regarding parents

and induction regarding peers. In all, there are a large number of significant findings especially for the major moral indexes—guilt and internalized moral judgments.

In contrast to the mothers, few significant findings were obtained for fathers—for boys as well as girls—and those that were obtained did not fit any apparent pattern.

A further step in the analysis of induction was to combine all indexes of this category into a composite index. The results, presented in Table 3, were quite striking in the case of mothers for all the moral indexes. Significant findings, all in the expected direction, were obtained for boys on guilt, internal moral judgments, confession, and acceptance of responsibility; and for girls on guilt, internal moral judgments, and consideration for others. When both sexes are combined, the findings are significant for all the moral indexes. The findings on identification are significant only for boys, however.

In contrast to induction, love withdrawal relates infrequently to the moral indexes (see Table 1). Further, in most cases in which significant relations between love withdrawal and moral development do occur, they prove to be negative. Taken as a whole, the importance of the distinction between love withdrawal and induction has been clearly demonstrated by these findings.

In sum it is a pattern of infrequent use of

⁵ Seven pages of tables giving medians for each of the high and low quartile groups have been deposited with the American Documentation Institute. Order Document No. 9079 from the ADI Auxiliary Publications Project, Library of Congress, Washington, D. C. 20540. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

TABLE 3

STATISTICALLY SIGNIFICANT RELATIONS BETWEEN
CHILD'S MORALITY INDEXES AND PARENT'S COM-
POSITE INDUCTION SCORE: MIDDLE CLASS

Morality index	Mother's induction			Father's induction		
	Boys	Girls	Sum	Boys	Girls	Sum
Guilt (child's response)	+*	+*	+*			
Internal moral judgment (child's response)	+*	+*	+*			
Confession (mother's report)	+**		+***			
Accepts responsibility (teacher's report)	+*		+*			
Consideration for other children (peers' ratings)		+**	+*			
Identification (child's response)	+*					

* $p < .05$.
 ** $p < .01$.
 *** $p < .005$.

power assertion and frequent use of induction by middle-class mothers which generally appears to facilitate the facets of morality included in this study.⁶

*The question might be raised here as to the extent to which these findings should be interpreted as independent. Do induction and power assertion exert independent influence on morality, or are they but two aspects of the same influence; for example, do the measures used require that someone high on induction is necessarily low on power assertion? The findings in Table 1 suggest the influences are largely independent. That is, there are only a few instances in which negative power assertion findings and positive induction findings for the same subsample were obtained with the same measure. In most cases the findings for the two types of discipline were obtained with different measures, and in some instances a finding was obtained for one but not the other

There is, however, one major exception to this pattern. The peers' reports of the boy's consideration for other children is positively related to the mother's report of their present use of power assertion (Table 1). A possible explanation of this finding is that our measure of consideration is a poor one especially for the boys. In particular, there is no built-in provision to assure that the behavior is based on internal motivation. The motive behind such behavior in the case of boys might instead often be a need for approval by peers. Why this should be the case for boys and not for girls remains unclear. It should be noted, however, that consideration is a more deviant value for boys than girls. Evidence for this is provided from a measure of values administered to the children. The largest sex difference found was on the consideration item ("goes out of his way to help others"). The girls valued this trait more than the boys ($p < .001$). Thus consideration does appear to have a different meaning for the two sexes.

Lower-class discipline. In discussing the lower-class findings the lack of parent interview data must be kept in mind. Nevertheless, there are several very apparent contrasts with the middle-class sample. Foremost among these is the general paucity of significant relationships between the child's moral development and his report of parental discipline. This is especially striking in the case of the mother's discipline. Furthermore, of

(e.g., guilt in boys relates to induction, but not to power assertion).

TABLE 4

STATISTICALLY SIGNIFICANT RELATIONS BETWEEN CHILD'S MORALITY INDEXES AND
MOTHER'S DISCIPLINE: LOWER CLASS

Morality index	Power assertion			Love withdrawal			Induction re parent		
	Boys	Girls	Sum	Boys	Girls	Sum	Boys	Girls	Sum
Guilt (child's response)				+c*				+c*	
Internal moral judgment (child's response)									
Accepts responsibility (teacher's report)								+c*	
Consideration for other children (peers' ratings)	+c*				-c*	-c*			
Identification (child's response)				c					

Note.—Interview data were not obtained from the lower-class parents. Thus all entries in Tables 3 and 4 are based on child reports. For the same reason lower-class data on confession and on induction regarding peers were unavailable.

* $p < .05$.

TABLE 5

STATISTICALLY SIGNIFICANT RELATIONS BETWEEN CHILD'S MORALITY INDEXES AND
FATHER'S DISCIPLINE: LOWER CLASS

Morality index	Power assertion			Love withdrawal			Induction re parent		
	Boys	Girls	Sum	Boys	Girls	Sum	Boys	Girls	Sum
Guilt (child's response)	-c*				+c*				
Internal moral judgment (child's response)									
Accepts responsibility (teacher's report)									
Consideration for other children (peers' ratings)									
Identification (child's response)	-c*		-c*				+c*		+c*

* $p < .05$.

those significant relationships that emerge, two are inconsistent with our expectations. First, as with the middle-class sample, the boy's consideration is related positively to the mother's use of power assertion. Second, in contrast with the findings for the middle-class boys, guilt is positively associated with the mother's use of love withdrawal, but unrelated to the mother's use of power assertion or induction. In summary, our expectations were not confirmed for the lower-class sample, and no general conclusion may be drawn.

The infrequent relationships between the child's moral development and the mother's discipline, compared to the middle-class sample, suggest that the lower-class mother's discipline may be less crucial and singular a variable. This in turn may be due to several factors. First, the mothers more often work full time in the lower than in the middle class. Second, the combination of large families and less space may result in the parent and child interacting with many other people besides each other. Third, according to the more traditional family structure usually found in the lower class (e.g., Bronfenbrenner, 1958), the father is more often the ultimate disciplining agent. In our sample, for example, boys more often reported that their mothers had the fathers do the disciplining ("says she'll tell your father") in the lower class than in the middle class ($p < .01$). Fourth, lower-class children are encouraged to spend more time outside the home than middle-class children. For all these reasons the socializing process may be more diffuse in the lower class; that

is, it may be more equally shared by the mother with the father, with siblings, members of the extended family, the child's peers, and others.⁷

Further research comparing the two classes needs to be performed. One might conjecture that because of the more diffuse socialization process in the lower class the basis of internalization may be quite different for children in the two classes, with consequent differences in the kind of morality that develops.

Affection. The relations between affection and the six moral indexes are presented in Table 6. The most notable features of this table are first, as expected, the relationships are positive; second, most of the findings, as with the discipline data, were obtained for middle-class mothers. It should also be noted that most of the findings are based on the child's report.

Role of the father. Several studies of delinquency (e.g., Glueck & Glueck, 1950; McCord & McCord, 1958; Miller, 1958) suggest that the father is important in the devel-

⁷ Another possible explanation for the paucity of findings in the lower class is that the lower-class children are very low on morality. Thus if the upper quartile of the lower class on morality were like the lower quartile of the middle class, there would be no reason to expect similar associations for the two classes. This possibility can be discounted since there was no overlap between the lower-class upper quartile and the middle-class lower quartile. And although there was a general tendency for the lower class to be lower on morality than the middle class, the difference was significant only for internal moral judgment and consideration for others, and only for girls.

opment of internal controls. Our findings, especially in the middle class, seem to suggest that this is not so. Relatively few significant relationships were obtained between paternal discipline and the child's morality, and several were in a direction opposite to that expected.

Of course, it is possible that the role of the father is more important than indicated in this study. For example, the father might provide the cognitive content of the standards by direct instruction rather than by his discipline techniques. Lacking data on direct instruction, we could not test this possibility. Another possibility is that the role of the father is a less direct one. That is, he may affect the moral development of the child by his relationship to the mother and his influence on the discipline techniques chosen by the mother. This is indicated in a study of preschool children where evidence was found suggesting that women who are treated power assertively by their husbands tend to react by using power assertive discipline on their children (Hoffman, 1963c). It may also be that the father's role is ordinarily latent in its effects and only becomes manifest under exceptional circumstances such as those often associated with delinquency. That is, under normal conditions with the father away work-

ing most of the time and the mother handling most of the disciplining, as in our middle-class sample, the father's importance may lie mainly in providing an adequate role model that operates in the background as a necessary supporting factor. Under these conditions, the specific lines along which the child's moral development proceeds may be determined primarily by the mother's discipline. An adequate role model is lacking, however, in extreme cases as when there is no father, when the father is a criminal, or when the father is at home but unemployed, and this may account for the findings obtained in the delinquency research.

Methodological issues. Any study of child rearing and moral development that relies on indexes of discipline and morality from the same source is open to the criticism that the relationships that emerge are due to the lack of independence of the sources. If that source is the child himself, the suspicion might be held that the child's report of parental discipline is simply another projective measure of the child's personality. It should be noted that in the present study the relationships between the child's morality and the parent's report of discipline were generally in the same direction as those involving the child's report of discipline. (We refer here to the middle-class-

TABLE 6
STATISTICALLY SIGNIFICANT RELATIONS BETWEEN CHILD'S MORALITY
INDEXES AND PARENT'S AFFECTION

Morality index	Middle class						Lower class					
	Mothers			Fathers			Mothers			Fathers		
	Boys	Girls	Sum	Boys	Girls	Sum	Boys	Girls	Sum	Boys	Girls	Sum
Guilt (child's response)	+c*		+c*			+n*						
Internal moral judgment (child's response)		+c*	+n*									
Confession (mother's report)	+c*		+c*			+p*						
Accepts responsibility (teacher's report)		+n*			+n*							
Consideration for other children (peers' ratings)	+p*	+c*	+c*				+c*			+c*		
Identification (child's response)	+c**	+c**	+c**			c	+c*					

* $p < .05$.

** $p < .01$.

mother findings.) In addition, over half the significant findings for each sex involve relations between measures obtained from different respondents.

Further support for our findings comes from a recent review in which our threefold discipline classification was applied to the previous research (Hoffman, in press). Since most studies used a power assertive-nonpower assertive dichotomy, as indicated earlier, the raw data were examined (and recoded where necessary) to determine whether love withdrawal, induction, or some other form of nonpower assertion was responsible for the findings. The results were clearly consistent with ours. Since a wide range of theoretical and methodological approaches were involved in the studies reviewed, our confidence in the findings reported here is considerably strengthened.

A common problem also relevant to the present design is that no definitive conclusion may be drawn about causal direction of the relationships obtained. Any solution to this will have to wait upon application of the experimental method or longitudinal studies. Nevertheless, some support for the proposition that discipline affects moral development, rather than the reverse, may be derived from the fact that several findings bear on the use of discipline in the past. If these reports are assumed to be reasonably valid, to argue that the child's moral development elicits different discipline patterns (rather than the reverse) necessitates the further assumption that the child's morality has not changed basically from early childhood. This is an unlikely assumption in view of common observations (e.g., about the child's changing acceptance of responsibility for transgression) and the findings about the developmental course of moral judgments obtained by Piaget (1948), Kohlberg (1963), and others.

Theoretical discussion. In this section we will analyze the disciplinary encounter into what we believe to be some of its most basic cognitive and emotional factors.

First, any disciplinary encounter generates a certain amount of anger in the child by preventing him from completing or repeating a motivated act. Power assertion is probably most likely to arouse intense anger in the

child because it frustrates not only the act but also the child's need for autonomy. It dramatically underscores the extent to which the child's freedom is circumscribed by the superior power and resources of the adult world. This is no doubt exacerbated by the fact that power assertion is likely to be applied abruptly with few explanations or compensations offered to the child. (The empirical evidence for a positive relation between power assertion and anger has been summarized by Becker, 1964.)

Second, a disciplinary technique also provides the child with (a) a model for discharging that anger, and may provide him with (b) an object against which to discharge his anger. The disciplinary act itself constitutes the model for discharging the anger which the child may imitate.

Third, as much animal and human learning research has now shown, what is learned will depend on the stimuli to which the organism is compelled to attend. Disciplinary techniques explicitly or implicitly provide such a focus. Both love withdrawal and power assertion direct the child to the consequences of his behavior for the actor, that is, for the child himself, and to the external agent producing these consequences. Induction, on the other hand, is more apt to focus the child's attention on the consequences of his actions for others, the parent, or some third party. This factor should be especially important in determining the content of the child's standards. That is, if transgressions are followed by induction, the child will learn that the important part of transgressions consists of the harm done to others.

Fourth, to be effective the technique must enlist already existing emotional and motivational tendencies within the child. One such resource is the child's need for love. This factor depends on the general affective state of the parent-child relationship, the importance of which may be seen in the consistent relationship obtained between affection and the moral indexes (Table 6). Given this affective relationship, some arousal of the need for love may be both necessary for and capable of motivating the child to give up his needs of the moment and attend to (and thus be influenced by) the parent's discipline tech-

equ. Too much arousal, however, may produce intense feelings of anxiety over loss of love which may disrupt the child's response especially to the cognitive elements of the technique. All three types of discipline communicate some parental disapproval and are thus capable of arousing the child's need for love. But it is possible that only inductions can arouse this need to an optimal degree because the threat of love withdrawal implicit in inductions is relatively mild. Also, it is embedded in the context of a technique which explicitly or implicitly suggests a means of reparation. Inductions are thus less likely to disrupt the child's response—as well as his general affective relationship with the parent—than either love withdrawal which may arouse undue anxiety, or power assertion which arouses anger and other, disruptive affects.

The second emotional resource, empathy, has long been overlooked by psychologists as a possibly important factor in socialization. Empathy has been observed in children to occur much before the child's moral controls are firmly established (e.g., Murphy, 1937). We believe that it is a potentially important emotional resource because it adds to the aroused need for love the pain which the child vicariously experiences from having harmed another, thus intensifying his motivation to learn moral rules and control his impulses. Of the three types of discipline under consideration, induction seems most capable of enlisting the child's natural proclivities for empathy in the struggle to control his impulses. As indicated in greater detail elsewhere (Hoffman, 1963b; Hoffman, in press; Hoffman & Saltzstein, 1960), we view induction as both directing the child's attention to the other person's pain, which should elicit an empathic response, and communicating to the child that he caused that pain. Without the latter, the child might respond empathically but dissociate himself from the causal act. The coalescence of empathy and the awareness of being the causal agent should produce a response having the necessary cognitive (self-critical) and affective properties of guilt.

It follows from this analysis that power assertion is least effective in promoting devel-

opment of moral standards and internalization of controls because it elicits intense hostility in the child and simultaneously provides him with a model for expressing that hostility outwardly and a relatively legitimate object against which to express it. It furthermore makes the child's need for love less salient and functions as an obstacle to the arousal of empathy. Finally, it sensitizes the child to the punitive responses of adult authorities, thus contributing to an externally focused moral orientation.

Induction not only avoids these deleterious effects of power assertion, but also is the technique most likely to optimally motivate the child to focus his attention on the harm done others as the salient aspect of his transgressions, and thus to help integrate his capacity for empathy with the knowledge of the human consequences of his own behavior. Repeated experiences of this kind should help sensitize the child to the human consequences of his behavior which may then come to stand out among the welter of emotional and other stimuli in the situation. The child is thus gradually enabled to pick out on his own, without help from others, the effects of his behavior, and to react with an internally based sense of guilt. Induction in sum should be the most facilitative form of discipline for building long-term controls which are independent of external sanctions, and the findings would seem to support this view.

Love withdrawal stands midway between the other two techniques in promoting internalization. It provides a more controlled form of aggression by the parent than power assertion, but less than induction. It employs the affectionate relationship between child and parent perhaps to a greater degree than the other two techniques, but in a way more likely than they to produce a disruptive anxiety response in the child. However, it falls short of induction in effectiveness by not including the cognitive material needed to heighten the child's awareness of wrongdoing and facilitate his learning to generalize accurately to other relevant situations, and by failing to capitalize on his capacity for empathy.

The weak and inconsistent findings for love withdrawal suggest that anxiety over loss of love may be a less important factor in the

child's internalization than formerly thought to be the case. Before drawing this conclusion, however, the possibility that love withdrawal is only effective when the parent also freely expresses affection, as suggested by Sears et al. (1957), should be considered. We were able to test this hypothesis by examining the relation between love withdrawal and the moral indexes within the group of subjects who were above and below the median on affection, and also within the upper and lower quartile groups. The results do not corroborate the hypothesis: the relations between love withdrawal and the moral indexes do not differ for the high- and low-affection groups.

In an earlier study with preschool children, however, love withdrawal was found to relate negatively to the expression of overt hostility in the nursery school (Hoffman, 1963b). It was possible to make a similar test in the present study since teacher ratings of overt hostility were available. Here, too, love withdrawal related negatively to hostility outside the home ($p < .05$).⁶ We also found that love withdrawal is used more when the child expresses hostility toward the parent than in other types of discipline situations. These findings suggest that the contribution of love withdrawal to moral development may be to attach anxiety directly to the child's hostile impulses, thus motivating him to keep them under control. Psychoanalytic theory may thus be correct after all in the importance assigned love withdrawal in the socialization of the child's impulses. Our data, however, do not support the psychoanalytic view that identification is a necessary mediating process. That is, we found no relation between love withdrawal and identification (Tables 1-4).

⁶ Power assertion related positively to hostility ($p < .05$), and induction showed a slight nonsignificant negative relation.

Some relevant experimental evidence is also available. Gordon and Cohn (1963) found that doll-play aggression expressed by children in response to frustration decreased after exposure to a story in which the central figure, a dog, searches unsuccessfully for friends with whom to play. Assuming the story arouses feelings of loneliness and anxiety over separation in the child—feelings akin to the emotional response to love-withdrawal techniques—these findings may be taken as further support for the notion that love withdrawal may contribute to the inhibition of hostility.

It remains possible, of course, that a form of unconscious identification which may not be tapped by our more consciously focused measure serves to mediate between the parent's love withdrawal and the child's inhibition of hostile impulses—as suggested in psychoanalytic theory.

In any case, our data do tend to show that love withdrawal alone is an insufficient basis for the development of those capacities—especially for guilt and moral judgment—which are critical characteristics of a fully developed conscience.⁷

⁷ It should be noted that love withdrawal might relate positively to guilt as defined in psychoanalytic terms, that is, as an irrational response to one's own impulses. Clearly our concept of guilt is quite different from the psychoanalytic, pertaining as it does to the real human consequences of one's actions.

REFERENCES

- ALLINSMITH, W., & GREENING, T. C. Guilt over anger as predicted from parental discipline: A study of superego development. *American Psychologist*, 1955, 10, 320. (Abstract)
- ARONFRED, J. The nature, variety, and social patterning of moral responses to transgression. *Journal of Abnormal and Social Psychology*, 1961, 63, 223-241.
- BECKER, W. Consequences of different kinds of parent discipline. In M. L. & L. W. Hoffman (Eds.), *Review of child development research*. Vol. 1. New York: Russell Sage Foundation, 1964. Pp. 169-208.
- BRONFENBRENNER, U. Socialization and social class through time and space. In E. E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1958. Pp. 400-425.
- GLUECK, S., & GLUECK, E. *Unraveling juvenile delinquency*. New York: Commonwealth Fund, 1950.
- GORDON, J. E., & COHN, F. Effect of fantasy arousal of affiliation drive on doll play aggression. *Journal of Abnormal and Social Psychology*, 1963, 66, 301-307.
- HILL, W. F. Learning theory and the acquisition of values. *Psychological Review*, 1960, 67, 317-331.
- HOFFMAN, M. L. Power assertion by the parent and its impact on the child. *Child Development*, 1960, 31, 129-143.
- HOFFMAN, M. L. Childrearing practices and moral development: Generalizations from empirical research. *Child Development*, 1963, 34, 295-318. (a)
- HOFFMAN, M. L. Parent discipline and the child's consideration for others. *Child Development*, 1963, 34, 573-588. (b)
- HOFFMAN, M. L. Personality, family structure, and social class as antecedents of parental power assertion. *Child Development*, 1963, 34, 869-884. (c)

- HOFFMAN, M. L. Socialization practices and the development of moral character. In M. L. Hoffman (Ed.), *Character development in the child*. Chicago: Aldine, in press.
- HOFFMAN, M. L., & SALTZSTEIN, H. D. Parent practices and the development of children's moral orientations. In W. E. Martin (Chm.), *Parent behavior and children's personality development: Current project research*. Symposium presented at American Psychological Association, Chicago, September 1, 1960.
- KOHLBERG, L. The development of children's orientations toward a moral order. *Vita Humana*, 1963, 6, 11-33.
- MCCORD, J., & MCCORD, W. The effect of parental role model on criminality. *Journal of Social Issues*, 1958, 14, 66-75.
- MILLER, W. Lower class culture as a generating milieu of gang delinquency. *Journal of Social Issues*, 1958, 14, 5-19.
- MURPHY, L. B. *Social behavior and child personality*. New York: Columbia University Press, 1937.
- PIAGET, J. *The moral judgment of the child*. Glencoe, Ill.: Free Press, 1948.
- SEARS, R. R., MACCOBY, E. E., & LEVIN, H. *Patterns of child rearing*. Evanston, Ill.: Row, Peterson, 1957.

(Received March 11, 1966)

A LEVEL OF ASPIRATION MODEL FOR GROUP DECISION MAKING¹

DONALD L. HARNETT

Graduate School of Business, Indiana University

Experimental and theoretical approaches to the concept of level of aspiration were used to develop a model which provides an equitable decision for a group faced with the problem of resolving individual differences into a collective choice. A series of experiments was designed to test alternative formulations derived from the level of aspiration model and from simple majority rule. Experimental results confirmed the hypothesis that under certain conditions the level of aspiration model would be able successfully to predict the decision alternative selected by an experimental group. The predictive efficiency of the model was shown to depend on the type of bargaining (isolated versus face to face), the nature of the conflict situation, and the decision rule assumed to prevail.

The research presented in this paper is primarily concerned with an experimental test of a descriptive model for studying group decision making, although the model is also conditionally normative. This model was designed to investigate the problem of finding a meaningful measure for aggregating individual preferences: one which avoids certain inconsistencies and the inadequacy of an ordinal scale (ranking) by considering strength of preference, yet one which does not encounter the interpersonal comparison problems common to an interval measure of utility.

The possibility of obtaining a meaningful measure of utility stronger than ordinality was suggested by von Neumann and Morgenstern (1947), whose numerical index of utility has the properties of both interval and ordinal scales. The implications of their research led Siegel (1956) to investigate the possibility of measuring utility on an ordered metric scale. Such a measure, which lies between an ordinal and an interval scale, not only ranks the alternatives, but ranks the relative distance between adjacent alternatives as well. This type of scale was shown to yield valuable data on individual values in Siegel's research on level of aspiration.

Siegel's hypotheses concerning level of aspiration were derived from assumptions simi-

lar to Simon's (1955) concerning the nature of individual utility functions.² That is, the utility of success and failure are assumed to be relatively large in relation to the difference between alternatives (goals), so that the utility function takes a large jump at the point separating success and failure. These assumptions, together with extensive theoretical and experimental research, led Siegel (1957, p. 257) to define level of aspiration as

... the least upper bound of that chord (connecting two goals) which has maximum slope; i.e., the level of aspiration is associated with the higher of the two goals between which the rate of the utility function is a maximum.

Under this definition, utility measured on an ordered metric scale is sufficient to determine an individual's aspiration level. Becker and Siegel (1958) collected data supporting this hypothesis by obtaining an ordered metric utility function based on subject responses to hypothetical choices between various alternatives. For example, with three alternatives a subject ranks the alternatives and then decides whether he would prefer his second choice for sure, or a 50-50 chance between his first and third choices. Should he choose the former (i.e., the second choice), this would indicate that the largest distance in utility is between Choice 2 and 3, and

¹The research reported in this paper was supported by a Ford Foundation grant. The author is especially indebted to Lawrence E. Fouraker and Harold Bierman who, among others, made valuable contributions to this research.

²Much of Siegel's (1956, 1957) research on level of aspiration, of course, stems from earlier work by Lewin, Dembo, Festinger, and Sears (1944). The relationship between the two is shown in Siegel's 1957 article.

that his second alternative is his aspiration level. Letting $U(X)$ represent the utility for alternative X , this result can be shown symbolically as follows:

$$\frac{1}{4}U(\text{1st choice}) + \frac{1}{2}U(\text{3rd choice}) < U(\text{2nd choice})$$

$$U(\text{2nd choice}) - U(\text{3rd choice}) > U(\text{1st choice}) - U(\text{2nd choice}).$$

For decision situations with more than three alternatives the subject is required to choose between probability combinations in a number of offers. If N is the number of alternatives to be scaled, then $\binom{N+1}{4}$ is the maximum number of probability combinations necessary; the exact number is usually considerably less and depends on the type of underlying ordered metric scale.

These developments concerning the interpretation and measurability of strength of preference on an ordered metric scale suggest a possible means of making interpersonal comparisons of utility and determining a criterion for "fair" group decisions—by comparing aspiration levels. A "fair" decision may be defined as one which satisfies more individuals than any other alternative, if one makes the assumption that group satisfaction depends only upon the number of individuals satisfied (i.e., reaching their level of aspiration).³ Each group member is weighted equally, according to whether or not the decision alternative in question meets his aspiration level. Let $U_i(A_j)$ be the utility of the j th alternative to the i th individual ($i = 1, 2, \dots, N$) and define a satisfaction index $S_i(A_j)$ as follows:

$$S_i(A_j) = \begin{cases} 1 & \text{if } U_i(A_j) \geq L_i \\ 0 & \text{if } U_i(A_j) < L_i \end{cases}$$

for all i and j . If L_i is the aspiration level of the i th person, then it is assumed group satisfaction will be largest for the alternative which

³ The measurement assumptions in the present model do not allow for individual differences in degree of satisfaction, although this factor may well be relevant for most group decisions.

maximizes M , where

$$M_j = \sum_{i=1}^N S_i(A_j).$$

By ordering all possible choices according to their value of M , a complete ranking of alternatives can also be obtained. Although max M is not necessarily unique, the additional information may be sufficient to eliminate possible intransitivities under simple majority rule.⁴

Because the level of aspiration model considers strength of preference, this model is hypothesized to provide, under certain circumstances, a better characterization of group decisions than will majority rule.

The level of aspiration solution is predicted whenever conditions tend to favor an equitable or "fair" resolution of conflict. Usually such conditions are present whenever a majority might be expected to give serious consideration to the preferences of a minority because of the desirability or necessity of compromise. Consider, for example, a group of three where one individual strongly prefers Alternative Y to Alternative X, and two individuals rate them nearly equal, but would rank X slightly ahead of Y. Simple majority rule dictates X as the group solution, even though all three individuals might agree that Y is the better group solution. The level of aspiration model, in contrast, takes individual strength of preference into consideration and would name Y as the group solution. A test of the predictive properties of the level of aspiration model relative to the simple majority rule, for the case of three individuals and three decisions alternatives, constitutes the empirical portion of the research presented in this paper.⁵

Experimental Design

In order to test the predictive ability of the level of aspiration model in relation to simple majority rule, a series of experiments was

⁴ See Arrow (1951) for an analytical treatment of the problem of intransitivity under simple majority rule.

⁵ The relationship between level of aspiration and group decision making also was investigated experimentally by Fouraker in 1964; his model and decision setting, however, differ substantially from those presented in this paper.

designed to simulate certain conflict settings involving three bargainers and three alternatives. These settings, or designs, had to represent combinations of individual preference arrays and aspiration levels such that the level of aspiration model predicted one solution while simple majority rule predicted another. An experimental test of the level of aspiration model was arranged by grouping three subjects and having them choose between three alternatives in certain specifically designed decision situations.

With three bargainers and three decision alternatives, there are 56 different preference arrays possible, only 26 of which may result

in possible discrepancies between the level of aspiration model and simple majority rule. If differences depending entirely on the labeling of alternatives are ignored, 6 different conflict situations remain from the 26. Two of these were eliminated from experimental consideration because they yield intransitive orderings under simple majority rule and thus do not provide a clear-cut alternative solution to the aspiration level model. Of the 4 remaining combinations, the 2 below (with their associated aspiration levels) were considered most representative of the conflict situation desired and were used as the basic design for the experimental tests.

Experimental Setting A

Individual			
	I	II	III
1	X	X	<u>Y</u>
Rank 2	<u>Y</u>	<u>Y</u>	X
3	<u>Z</u>	<u>Z</u>	Z
Majority rule solution			
			X
Level of aspiration solution			
			Y

Experimental Setting B

Individual			
	I	II	III
1	X	<u>Y</u>	<u>Z</u> *
Rank 2	<u>Y</u>	X	X — = level of aspiration
3	<u>Z</u>	Z	Y
Majority rule solution			
		X	
Level of aspiration solution			
		Y	

The nature of the conflict situation, the type of decision rule imposed, and the amount of interpersonal contact were hypothesized to influence support for one of the two solutions. A tendency toward a majority solution should be quite strong in Experimental Setting A, since the majority solution is readily identifiable and accessible. Experimental Setting B, on the other hand, has no initial "obvious" solution, and therefore a greater tendency toward the level of aspiration alternative could be expected. The nature of the decision rule assumed to prevail (e.g., majority or unanimous rule) should also play an important role in determining group decisions—the larger the number required to agree, the greater will be the tendency toward a compromise solution. Finally, variations in interpersonal contact are expected to influence

group outcomes due to an expected tendency toward "fairness" as increased interaction forces group members to defend their actions publicly. The other extreme, where the identity of bargaining opponents is kept anonymous (e.g., decisions are made by ballot), should minimize this tendency toward fairness.

Experimental tests were designed to introduce four of the eight possible combinations of these variables. The level of aspiration solution was expected to hold when at least one of the three variables favored such a solution. Arrangement of the experiments was thus in four designs, with the following variables and hypothesized solutions. (Variables with asterisks are assumed to favor the level of aspiration solution.)

Design	Variables			
	Experimental setting	Decision rule	Interpersonal contact	Hypothesized solution
I	A	Majority	Isolated bargaining (voting)	Majority rule
II	A	Majority	Face-to-face* bargaining	Level of aspiration
III	A	Unanimous*	Isolated bargaining (voting)	Level of aspiration
IV	B*	Majority	Isolated bargaining (voting)	Level of aspiration

Although a majority of cases are predicted to support the level of aspiration solution, some differences in the degree of agreement are expected—strength of support is hypothesized to increase from Design I to Design IV. Weakest support is predicted for Design I because this particular combination does not include any variables expected to favor the level of aspiration solution. The majority rule solution should hold in this case because two individuals are expected to impose their will on the third, the mere existence of a decision “fair” to all three persons being considered insufficient to induce either one to yield from his optimal choice. It would not be irrational to expect the level of aspiration model to hold, however, as the positive utility associated with the behavioral alternative of voting for a “fair” solution may bring the net utility for this alternative high enough to induce either Individual I or II (or both) to vote for it.⁶

Face-to-face bargaining in the identical situation should increase the tendency toward the level of aspiration solution, as the third party now has an opportunity to confront his opponents with his preferences and “appeal” for fairness. Hypothesized to aid this appeal and provide support for the aspiration solution is an expected natural reluctance for individual group members openly to reject what might be considered a fair solution. Needless to say, the assumption of such a natural reluctance is somewhat tenuous, especially in a competi-

tive society such as ours, so that strong support for the aspiration model is doubtful.

The unanimity requirement in Design III should provide a stronger support for the level of aspiration model, because now Individuals I and II cannot impose their will on the third party. Realizing that Individual III is not likely to accept what he considers an unfair solution, Individuals I and II should concede after considering the probable consequences of a refusal to concede (a stalemate) and in view of the fact that a concession decision will still satisfy their aspiration levels. Nevertheless, the fact that two individuals must give in to a third still runs contrary to what one would expect in a decision situation, and thus weakens expected support for the level of aspiration solution and increases the possibility of a simple majority rule solution or a stalemate. The implicit possibility of a “stalemate alternative” (i.e., a status quo solution) will be eliminated in the present research by defining a lack of consensus under the stated decision rule to represent a decision for each individual’s *worst* alternative.

Strongest support for the level of aspiration solution is predicted for Design IV. Here the setting is such that no initial majority rule solution exists, and at least one individual must vote for his second choice for the group to come to a decision. If one group member can yield to his second choice in this circumstance and still be satisfied with the outcome, clearly it will be to his advantage to do so rather than risk the possibility of a stalemate.

⁶ Individuals I, II, III correspond to the individuals whose preference arrays comprise Experimental Settings A and B.

Experimental Methods

For each of the experiments representing the four designs described above, the appropriate conflict characteristics were obtained by subject placements. These subjects were chosen because of their exhibited preferences for decision alternatives corresponding to those needed experimentally. The task for each group was to decide on the type of multiple-choice exam, out of three specified, they would prefer to take. Individual incentive was provided each participant by paying him in relation to his performance on the exam decided upon by the group. All three individuals had to take the same type of exam, and each person was paid (\$.50) for each question he answered correctly on this (eight-question) exam. Thus, competition was achieved by composing each group with individuals of different backgrounds and then offering them an appropriate array of exams from which to choose.

Arrangement of the volunteers into decision groups was determined in advance of the decision sessions by means of a questionnaire. This questionnaire asked each volunteer to rate his expected performance on an exam in each of 14 different areas by judging his performance in relation to a person with three or four college courses in that area. On the basis of these evaluations of expected performance, it was possible to arrange groups with a high likelihood of simulating the appropriate decision setting. Actual rankings and aspiration levels were obtained during the experimental session itself, so that this procedure produced only an approximation of the desired setting. It was therefore decided beforehand to consider as acceptable data only the decisions resulting from those groups where *actual* preferences and aspiration levels corresponded to the desired experimental setting.

The nature of the experiments for Design I, III, and IV dictated the testing of a large number of decision groups simultaneously, primarily to maintain the anonymity of individual group members. With a large number of subjects, group membership was kept secret by giving an individual information concerning his opponents' preferences while revealing only the fact that he was bargaining with other students in the same room. This was accomplished by using information sheets especially constructed for each experimental design and filled out for each participant in advance of his participation. These sheets described the nature of the experiments, listed the decision alternatives, and provided for each group member a brief summary of the preference strengths of his bargaining opponents.⁷ Statements of the following type were used in describing preference strengths (Student 2 refers to a bargaining opponent):

⁷ Since these sheets were all prepared in advance of each session, the preference strength attributed to each individual's opponents was an expected one. Only when these expectations were upheld were the data considered acceptable.

Student 2 ranks the three exams in the following order:

1st choice (Best performance)	German
2nd choice (Next best performance)	mathematics
3rd choice (Worst performance)	chemistry

Student 2 would strongly prefer to take the examination in German, and of the three exams, he would expect satisfactory performance *only* on that one type. Although he does prefer mathematics to chemistry, he would expect an unsatisfactory score in both areas.

The information sheets asked each participant to list his own ranking of the alternatives, and to decide whether he would prefer having a 50-50 chance between taking his first- or third-ranked exams, or taking his second-ranked exam for sure. This part of the information sheet determined each person's actual ranking and aspiration level and was used to decide whether or not a particular group conformed to the desired setting.

Each group under isolated bargaining conditions (voting) was given five votes to reach a decision, with the outcome of each ballot being made known before the next vote was called for. If at the end of five votes a decision had not been reached, each participant had to take the one examination, of the three specified, which received his worst rating on the original questionnaire.

At the beginning of each group session a brief oral description of the nature of the experiment was given to all participants. It was impressed upon the subjects that they would make their decision by voting, and that their votes would be their only means for expressing their desires. Each participant was then given his instruction sheet; after sufficient time had been allowed for subjects to read this material and to fill out the required information, the balloting was started. When all groups had reached a decision (or after five votes), the proper exams were distributed. Students were free to leave as soon as they had finished their exam and had been paid.

Design II required separate experimental sessions for each group in order to facilitate observation of the face-to-face bargaining process. Procedures for these groups were similar to those described above. In the introductory remarks it was emphasized that participants could make their decision in any manner they wished, and that it need not be unanimous. Subjects were then given their information sheets, and when these had been filled out and collected the participants were instructed to begin discussions leading toward their decision.

Discussion of Results

Although 144 students were selected to participate in the experiments, only 42 groups (126 subjects) were formed because a number of students failed to appear. An additional

loss of 30 subjects (10 groups) resulted from unanticipated aspiration levels or rankings by 11 subjects. Eight of the 11 changed away from their expected ranking, and 3 away from their anticipated aspiration levels. All 3 of the latter changes resulted from an increase in aspiration levels, that is, from an expected aspiration level of Choice 2 to an actual aspiration level for Choice 1. The 8 changes from expected rankings were all by individuals with an aspiration level for their second choice and in each case involved an interchange of Choices 1 and 2.

Stalemate solutions resulted from two groups in Design III and one group in Design IV. These solutions were difficult to classify in terms of the research hypothesis; this behavior does not conform to the predicted simple majority rule solution, nor is clear support indicated for the level of aspiration solution. Since the model was designed to handle only those groups reaching a decision, stalemates will be counted as evidence neither for nor against the research hypothesis. After subtracting stalemates, there were 7 acceptable decision groups in Design I, 9 in Design II, 7 in Design III, and 6 in Design IV.

Design I (Setting A, majority rule, isolated bargaining). Support for the predicted (simple majority rule) solution in Design I turned out to be stronger than originally assumed. Although strong support was anticipated, it had been expected that at least one or two individuals would choose the level of aspiration alternative because of its "fairness." As it was, every participant acted entirely in view of his own interests by voting for his first-ranked alternative, the result being a majority rule decision for all groups on the first ballot. A binomial test, with $P = Q = \frac{1}{2}$, was used to test the research hypothesis that experimental decisions would significantly favor the simple majority rule solution. The fact that all seven allowable groups made decisions consistent with simple majority rule permits rejection of the null hypothesis at $p = .008$ in favor of the research hypothesis.

In order to attain some idea of whether fairness was generally associated with the level of aspiration solution, each participant in Design I was asked (after the vote) to

write on the back of his instruction sheet which examination he considered a "fair" solution for his group. A majority of subjects indicated they considered the level of aspiration solution as the fair decision. Especially significant, however, is the breakdown of evaluations between the two persons in the dominant position (Individuals I and II) and the person in the weak position (Individual III). Alternatives listed as fair by the 27 participants who had the proper ranking and aspiration level are shown in Table 1.

Thus, the position each person assumed in the conflict situation seemed to play an important role in determining what he considered a fair solution, but not in determining his vote. Note, in this regard, that 16 out of 27 participants (10 individuals in Positions I and II, and 6 in Position III) did not vote for the alternative they listed as fair. Most surprising is the fact that 6 of the 10 persons in the weak positions (Individual III) listed simple majority rule as the fair solution. These subjects usually stated, in effect, that it would be unfair for them to impose their will on two other persons, and that majority rule by itself is a fair decision method. This itself would not be quite as surprising if an approximately equal percentage of dominant individuals held a similar view. The fact that only 5 out of 15 agreed, however, seems to indicate a basic dissimilarity between the two groups.

Design II (Setting A, majority rule, face-to-face bargaining). Four groups decided on the level of aspiration solution, and five decided on the examination supporting simple majority rule in Design II. A binomial test on this data does not permit rejection of the null hypothesis in favor of the research hypothesis that decisions will significantly

TABLE 1
NUMBER OF PERSONS LISTING THE ALTERNATIVE
SOLUTIONS AS "FAIR" IN DESIGN I

Fair solution	Individual		
	I, II	III	Total
Level of aspiration	10	4	14
Simple majority rule	5	6	11
Both of the above	2	0	2

support the level of aspiration solution ($p = .746$).

Much of the support for the aspiration solution was hypothesized to result from Individual III's appeal for "fairness," but if Individual III considers simple majority rule fair, then this support is lost. In fact, group decisions at times did seem to depend largely on Individual III's evaluation of fairness, and Individual I or II (or both) often indicated some willingness to compromise. In view of the results of Design I, it therefore is surprising that the level of aspiration model received as much support in Design II as it did.

With only a few exceptions, the participants usually asked each other about their ranking, major in college, and sometimes the number of courses taken in the subject area in question. During the course of the discussions most persons offered an accurate representation of their preferences. (E.g., "I would prefer physics, but wouldn't mind economics.") A few, perhaps wisely, were reluctant to admit they had knowledge in more than one subject area even when their aspiration level was for Choice 2.^a

Design III (Setting A, unanimous rule, isolated bargaining). Six of the nine group decisions in Design III supported the level of aspiration model, one supported simple majority rule, and two ended in stalemates. The binomial test allows for rejection of the null hypothesis at $p = .062$ in favor of the research hypothesis that decisions will significantly favor the level of aspiration solution.

As expected, Design III provided support for the level of aspiration model and also resulted in the largest number of stalemates. Not anticipated, however, was the high degree of reluctance for Individuals I and II to yield to their second choice although in retrospect this does correspond to the results of Design I. In three of the six groups supporting the aspiration model all five votes were necessary before a decision could be reached.

^a If any subject mentioned the possibility of side payments, an announcement was made that this would not be allowed. The possibility of implicit cooperation between friends was minimized by pairing volunteers who had little chance of knowing each other.

Only 2 of the 12 individuals who gave in to the aspiration solution did so on their first vote and only 2 on their second vote. The average individual in this set yielded in 3.16 votes, while it took, on the average, 4.00 votes for each of these groups to come to a decision. Results similar to these were obtained in a replication of Design III under slightly different conditions: out of six groups, there were four level of aspiration solutions and two stalemates. In this case, however, the two groups which were stalemated were allowed to meet and discuss their decision. Although no side payments were permitted (the payoff was for "extra" points toward their final grade), both groups were able to resolve their differences and unanimously agreed on the level of aspiration solution.

Design IV (Setting B, majority rule, isolated bargaining). Strongest support for the level of aspiration model was obtained in Design IV. Six of the seven groups decided on the examination supporting the level of aspiration model, the seventh ending in a stalemate. The results of this design permit rejection of the null hypothesis in favor of the research hypothesis that decisions will significantly support the level of aspiration solution. The binomial test allows for rejection at $p = .016$ when stalemates are ignored.

The average number of votes necessary in this design for those groups coming to a decision was 3.14, and each decision took at least two votes. Participants appeared to recognize readily the essential nature of the conflict situation, although few seemed to note the existence of a simple majority rule solution. Most subjects interviewed believed the outcome depended primarily on who was willing to give in first, and that Individual I was the most likely candidate. One person in the role of Individual I waited until the fifth vote to change his vote. He remarked, quite logically, that he had nothing to lose by waiting until the last moment since he could have gained if either one of his opponents had decided to change his vote.

A number of interviews and a replication of this design indicated the importance of accurately assessing Individual III's aspiration level. For if Individual III foresees that he has little or no chance of gaining his first-

TABLE 2
DECISION RESULTS FOR DESIGNS I-IV

Design	Level of aspiration solution	Majority rule solution	Stale-mates	Total	Level of significance for hypothesized solution
I	0	7	0	7	.008
II	4	5	0	9	.746
III	6	1	2	9	.062
IV	6	0	1	7	.016

ranked alternative, and if he has any incentive to actively pursue his second alternative over his third, it would certainly be worth his while to do so, rather than hold out for Choice I. Indeed, this is exactly what took place in the above-mentioned replication, the result being five out of six decisions for simple majority rule and only one for the level of aspiration solution. Of course, it was important in this latter study that Individual I recognized the weakness of III's position and was not in any hurry to change his own vote.

Summary

Table 2 shows the complete results for all four designs, together with the significance level for the hypothesized solution. The level of significance between designs, using a Fisher exact-probability test, is shown in Table 3.

As can be seen from the data in Tables 2 and 3, the hypothesized increase in strength of support for the level of aspiration solution from Design I to Design IV was for the most part upheld. Only the difference between Designs III and IV stands out in contrast to the predicted degree of difference, the deviation in this case due to the fact that both designs offered support for the model.

Conclusions

In view of the experimental results, it seems clear that the equity associated with the level of aspiration solution played an important

part in determining most group decisions. The strength and nature of this "fairness," which appeared to depend largely on the nature of the conflict situation, influenced decisions, in general, in one of two ways.

The first way, identified with Designs I and II, involved the "fairness" assumed to be inherent in a compromise solution. That is, a solution satisfying all participants was predicted to have special appeal even to group members who by virtue of their superior bargaining position could insist on a more favorable outcome. Although this aspect of fairness, as expected, did not find significant support under conditions of isolated bargaining and majority voting (Design I), the surprising fact was that not even one group decision supported the level of aspiration alternative.

Some support for the aspiration solution was found in Design II, indicating that face-to-face bargaining does tend to promote this type of fairness. In general, however, evidence supporting the importance of fairness of the sort where individuals voluntarily yield to a compromise solution was quite limited. Most participants based their decisions on their own personal rewards, thus favoring the majority rule solution. Simple majority rule was found to have a measure of fairness all its own, although this also depended on the role assumed.

The second and strongest influence of "fairness" resulted from Designs III and IV, where the combination of variables was such that no two individuals were able to insist on their first choice. The notion of fairness in these experiments seemed to give participants an opportunity to evaluate their chances of receiving the alternative they ranked first. Each person was able to assess his chances by noting the preferences of his opponents and considering whether it would be "reasonable" or "fair" to expect either one or both to give in. (E.g., "Would I yield in a comparable situation?") Most individuals would not yield to an unsatisfactory solution, nor would they expect their opponent(s) to do so either. Thus, the participant forced with choosing between a satisfactory second-ranked alternative or the possibility of a stalemate would be expected to yield. The importance of this

TABLE 3
LEVEL OF SIGNIFICANCE OF DIFFERENCE
BETWEEN DESIGNS

II	.069		
III	.002	.121	
IV	.004	.042	.539
Design	I	II	III

aspect of fairness was substantiated by the support given the level of aspiration solution in Designs III and IV.

In general, the following conclusions were obtained from analysis of the data:

1. The mere existence of a solution fair to all was not sufficient to induce behavior toward the level of aspiration solution. Majority rule itself had some connotation of fairness and provided a stable outcome under isolated bargaining conditions.

2. A prominent majority rule solution appeared to offset some of the strength of the level of aspiration alternative because participants could more readily identify and attain this solution. When no initial majority rule solution was obvious, there was a strong tendency for subjects to search for a means to determine a "reasonable" solution, which in this case was the level of aspiration alternative.

3. The decision rule assumed to hold had a significant effect on the decision alternative selected. Majority rule lends support for the simple majority rule solution, while requiring unanimity favors the level of aspiration solution.

4. Increasing the amount of interpersonal contact increased the number of "fair" solutions, but did not result in significant support for the level of aspiration alternative.

REFERENCES

- ARROW, J. *Social choice and individual values*. New York: Wiley, 1951.
- BECKER, S. W., & SIEGEL, S. Utility of grades: Level of aspiration in a decision theory context. *Journal of Experimental Psychology*, 1958, **55**, 81-85.
- FOURAKER, E. Level of aspiration and group decision making. In S. Messick & A. H. Brayfield (Eds.), *Decision and choice*. New York: McGraw-Hill, 1964. Pp. 201-239.
- LEWIN, K., DEMBO, T., FESTINGER, L., & SEARS, P. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavioral disorders*. Vol. 1. New York: Ronald Press, 1944. Pp. 333-378.
- NEUMANN, J. VON, & MORGENTHAU, O. *Theory of games and economic behavior*. (2nd ed.) Princeton: Princeton University Press, 1947.
- SIEGEL, S. A method for obtaining an ordered metric scale. *Psychometrika*, 1956, **21**, 207-216.
- SIEGEL, S. Level of aspiration and decision making. *Psychological Review*, 1957, **64**, 253-263.
- SIMON, H. A. A behavioral model of rational choice. *Quarterly Journal of Economics*, 1955, **99**, 99-118.

(Received March 13, 1966)

FAMILY ENVIRONMENT AND THE DEVELOPMENT OF DEFENSE AND COPING MECHANISMS¹

ALLAN R. WEINSTOCK

University of Wisconsin Medical School

In order to study the effects of family environment on the development of ego mechanisms, 39 males who had been followed longitudinally were used as Ss. Ratings of the parents' personalities and of Ss' childhood family environments were correlated with ratings of Ss' use of 10 defense and 10 coping mechanisms at 30 years of age. Results indicate that primitive defenses (denial and repression) are related to the father's passivity in the early family environment. More differentiated defenses are related to rejection of S in early adolescence. Expressive coping (tolerance of ambiguity, and regression in the service of the ego) is correlated with family conflict in adolescence. In general, Ss' imitation of parental behavior plays an important role in the development of particular defense and coping mechanisms.

Several theorists have suggested that individual differences in the use of defense mechanisms and adaptive techniques are closely related to the family environment experienced in childhood and remain as relatively fixed aspects of the adult character structure. However, there have been few studies of the childhood antecedents of defense mechanisms. Miller and Swanson (1960) have found that denial is related to severe discipline, unexplained requests by the parents, and few rewards. Repression was found significantly more in middle-class families which use psychological discipline than in working-class families which use corporal discipline. In a study of mother-child interactions in a problem-solving situation with high and low defensive boys, Ruebush, Byrum, and Farnham (1963) found that low defensive males have mothers who are more approving, encouraging, and warmly affectionate, while high defensive males have mothers who are less supporting and accepting of their sons and less responsive to their needs.

Psychoanalytically oriented theorists have tended to discuss the establishment of these mechanisms in terms of complex reactions

within the child and have focused on intrapsychic conflict in determining defensive choice. Studies which have focused on the relationship between intrapsychic conflict and the development of specific defense mechanisms have yielded few significant results. In Miller and Swanson's (1960) research the severity of weaning and bowel training had little direct relationship to the establishment of particular mechanisms. However, variables describing the social situation seemed to influence the adoption of particular defenses. Swanson (1961) pointed out that intrapsychic conflict is only a minimal condition necessary for the establishment of defensiveness and could lead to a number of alternative defenses. He suggested that imitation plays an important role in determining the defense mechanism chosen. Hetherington and Brackbill (1963) have similarly found that identification with the like-sex or dominant parent is more strongly related to anal character traits than experiences related to bowel training.

The purpose of the present investigation is to explore the relationships between a large number of variables describing the childhood family environment and ratings of several defense and coping mechanisms in order to delineate the family antecedents of specific ego mechanisms. More generally, this research is directed at clarifying the role of imitative behavior in the development of ego mechanisms and eliciting general principles about

¹The subjects for this study are from the University of California Guidance Study. The author wishes to thank the Guidance Study staff, Norma Haan, and T. R. Sarbin for their assistance in carrying out this investigation. This research was supported by National Institute of Mental Health Grant 5-F1 MH-23, 528-02.

the way in which family environment influences character formation.

METHOD

Sample

The subjects are participants in the University of California Guidance Study. These subjects were drawn from the Berkeley Survey which included every third child born in Berkeley between January 1, 1928 and June 30, 1929. The psychological, mental, and physical development of the subjects was studied intensively at 21 months of age. Home visits were made, and office interviews with the child, his parents, and siblings were made at 6-month or yearly intervals until the subject was 18 years old. Family members were encouraged to discuss their problems and concerns freely and to contact the Institute for guidance when problems arose. The guidance group, a subsample of the Guidance Study, originally consisted of 65 male and 61 female subjects and their families. Thirty-nine of the male subjects returned for intensive interviews when they were 30 years old. These 39 male subjects comprise the sample used in the present study.²

Assessment Procedure

Early family environment. A series of intensive home and office interviews were conducted with both parents when the subject was between 21 and 36 months old. The clinical director interviewed the parents during office visits and collected information on the family environment and marital situation. Family problems were frequently discussed during these visits. A psychiatric social worker visited the home, took a detailed history on the habits and regimen of the subjects, and observed parent-child relationships. Sixty-two variables describing the family situation were rated by each of the two case-workers. These included ratings of the parents' personalities, their behavior toward the subject, and their relationship toward each other. These ratings were made on 5-point scales. The interrater reliabilities for these scales range from .38 to .72. Since the two raters used different sources of information, these coefficients may be more properly considered as measures of the validity of the ratings. A system of weightings was worked out to take account of discrepancies in the amount of information obtained on an item by each rater. The ratings of the two interviewers were composited, and these composite ratings were used in the present study (MacFarlane, 1938).

Family environment during early adolescence. In order to investigate the extent to which early as compared to later family environment influences the development of particular defense and coping mechanisms, ratings of the family during the subjects' early adolescence were also used in the present study.

The subjects' parents were interviewed yearly when the subjects were between 5 and 16 years old. Ratings of personality traits of the parents, the family environment, discipline, and the relationship between the subject and his parents were made each year by the interviewer on 5-point scales. Since the ratings were incomplete, and since they were made by only one interviewer, the yearly ratings were composited into four 3-year periods (5-7, 6-10, 11-13, 14-16). Ratings of the period 11-13 years were used in the present study.

The families of 33 of the 39 subjects who participated in the adult follow-up interviews were assessed when the subjects were between 11 and 13 years old. However, the number of subjects rated on a particular variable is between 6 and 33. Forty-two family variables were rated during this period.

Ego mechanisms. Intensive office interviews lasting as long as 16 hours were conducted with the subjects when they were approximately 30 years old. Three interviews were conducted with each subject. In the first interview the subject was asked to elaborate on the high and low points of his life and to discuss his experiences since his last formal visit to the Institute at adolescence. In the second interview the subject was asked to describe his parents and to compare them to each other and to himself. The interview was structured to provide information on personality, physical characteristics, reactions to stress, interests, and social interaction. A third interview was conducted with subjects who were or had been married. It was similar in structure to the second, but compared the subject to his wife and her family.

Detailed notes were taken during the interviews and transcribed and typed after completion of the interview. The interview material was read by two clinical psychologists experienced in clinical assessment and in rating case material. The raters, working independently, read through the entire case material on a subject and rated the subject for his use of each of the 20 defense and coping mechanisms previously defined by Haan (1963) and Kroeber (1963). Each mechanism was rated on an 8-point continuous scale. A number of logically consistent relationships between ratings obtained by this procedure and other variables, including intelligence, social mobility, Rorschach, and MMPI scores attest to the validity and usefulness of the rating procedure (Haan, 1963, 1964a, 1964b; Kroeber, 1963).

The reliabilities of the two raters for the 39 cases used in the present study are reported in Table 1. The mean of the scores assigned to a variable by the two raters was taken as the subjects' score in all further data analysis. Interrater reliabilities for rationalization, reaction formation, and objectivity did not approach a significance level of .05 and were defined as unreliable.

Analysis

Since relatively few studies of the family antecedents of ego mechanisms have been carried out, the present research was conceived as an exploratory

² A more comprehensive description of the sample and the childhood assessment procedures may be found in MacFarlane (1938).

TABLE 1
RELIABILITIES OF EGO MECHANISMS

Defense mechanism	Reliability ^a	Coping mechanism	Reliability ^b
Isolation	.61	Objectivity ^a	.40
Intellectualizing	.80	Intellectuality	.77
Rationalization ^a	.43	Logical analysis	.73
Doubt	.82	Tolerance of ambiguity	.68
Denial	.65	Concentration	.73
Projection	.63	Empathy	.72
Regression	.75	Regression in the service of the ego	.67
Displacement	.53	Sublimation	.68
Reaction formation ^a	.37	Substitution	.72
Repression	.77	Suppression	.65
Total defense	.80	Total coping	.78

^a Unreliable.

^b Reliabilities were corrected for attenuation by the Spearman-Brown "prophecy" formula.

tory investigation, and all variables describing the childhood family environment of the subject were included in the data analysis. The 62 early family variables and 42 adolescent variables were correlated with each of the 17 reliably rated ego mechanisms. In order to reduce the data to comprehensible form, only those variables which are significantly correlated with an ego mechanism at $p < .05$ (two-tailed test) are reported.³ Twenty-four variables describing the early family environment and 13 adolescent variables had no significant linear correlation with any ego mechanism.

RESULTS AND DISCUSSION

Defense Mechanisms

Denial. Denial is highly related to several characteristics of the father in early childhood; the father may be characterized as calm, friendly toward members of the family, and indifferent toward environmental stresses. The few correlations which involve the mother suggest that she is also indifferent and agreeable. A negative correlation between denial and social class for these families (Weinstock, 1965) would lend further support to the conclusion that the families of subjects who use denial are indifferent to the environmental stress and model the behaviors associated with denial. Byrne (1964) has also found that subjects using denying and repressive defenses characterize their families as permissive, accepting, and warm. However,

these results are in sharp contrast to Miller and Swanson's hypothesis that denial is related to a harsh early environment.

The findings summarized in Table 2 indicate that repression is related to many of the same family variables which are associated with denial. However, a greater emphasis on the father's avoidance of conflict and the absence of association between repression and social class (Weinstock 1965) suggest that the father's indifference and withdrawal may reflect an avoidance of impulses rather than an avoidance of the environment. The father's modeling of behaviors associated with repression is again related to the subject's use of this defense.

Both repression and denial are highly related to the early family environment, but only slightly related to the family environment at adolescence. These findings offer tentative confirmation to Miller and Swanson's (1960) hypothesis that denial and repression are primitive mechanisms which are established early in the child's development when his cognitive structure is relatively undifferentiated.

TABLE 2

SIGNIFICANT CORRELATIONS BETWEEN PRIMITIVE DEFENSES AND FAMILY ENVIRONMENT

Denial		Repression	
Variable	r	Variable	r
Early family environment ^a		Early family environment ^a	
Father even-tempered	.46**	Agreement on expenditures	.37*
Father relaxed, indifferent	.40*	Mother adjusted to working or not working	.37*
Mother indifferent toward health	.38*	Father withdraws in conflict	.26*
Father concerned with liberty vs. commitment	.35*	Father relaxed, indifferent	.36*
Agreement on expenditures	.34*	Father even-tempered	.34*
Mother adjusted to working or not working	.34*		
Friendly, easy relationship—father towards mother	.33*		
Satisfaction over size and management of income	.32*		
Friendliness—father towards child	.32*		
Family environment at adolescence ^b			
Father relaxed, indifferent	.41*		

^a N = 39.

^b N = 27.

* $p < .05$.

** $p < .01$.

³ A complete listing of the family variables together with their definitions may be found in Weinstock (1965).

Regression and doubt. Regression and doubt which have in common the subject's inability to assert himself and commit himself to independent decisions are highly inter-correlated ($r = .61$) in the present study and are related to a large number of similar family variables both in early childhood and in early adolescence. Both mechanisms are correlated with the father's poor health and inactivity, withdrawal in conflict and general inadequacy, with the mother's psychological instability, and with poor adjustment between the parents in early childhood. In early adolescence both variables are related to the mother's brutality (i.e., attacks on the child's self-esteem) and poor relationship with her child, as well as strain and conflict in the home.

The development of these defenses is easily explained from a social learning perspective. For both mechanisms the parents seem to model feelings of doubt, inadequacy, and immature behavior for the child. The child grows up in an environment where conflicts

and problems are met with feelings of inadequacy and passivity. He has little opportunity in the family environment to observe active mastery and coping. Furthermore, it is probable that the parents' rejection and neglect of the subject lead to feelings of his own inadequacy and worthlessness.

Displacement. Several studies of the displacement of aggression suggest that this defense is related to the parents' restrictiveness and with the parents' modeling of aggressive behavior (Bandura & Walters, 1963). While displacement is significantly related to few family variables in the present study, those significant relationships which were found are in agreement with previous findings that the displacement of impulses is related to a restrictive family environment.

Obsessional defenses. In the present study intellectualization was not related to an interpretable group of family variables. The finding that this mechanism is highly correlated with intelligence (Haan, 1963) and with

TABLE 3
SIGNIFICANT CORRELATIONS BETWEEN REGRESSIVE DEFENSES AND FAMILY ENVIRONMENT

Regression			Doubt and indecision		
Variable	r	N	Variable	r	N
Early family environment		39	Early family environment		39
Father withdraws in conflict	.41**		Poor health—father	.50**	
Mother psychologically unstable	.40*		Poor sex adjustment	.47**	
Poor health—father	.40*		Mother psychologically unstable	.42**	
Mother unwilling to give sex instructions	.39*		Father's poor physical stamina	.42*	
Low income	.39*		Father concerned over having more children	.37*	
Father physically weak	.36*		Father tense or worrisome	.37*	
Conflict over size and management of income	.33*		Poor marital adjustment	.36*	
Family environment at adolescence			Poor health—mother	.36*	
Straining relationship with child—mother	.53**	31	Father withdraws in conflict	.36*	
Mother brutal	.51**	30	Family environment at adolescence		
Unwise discipline—mother	.76*	8	Mother brutal	.64***	30
Father worrisome	.44*	27	Strain and conflict in the home	.90**	8
Mother psychologically unstable	.42*	30	Poor discipline	.90**	6
Father restrictive	.41*	29	Unsatisfactory play—father with child	.53**	32
Conflict over discipline	.41*	33	Straining relationship—mother with child	.51**	31
Child disliked by mother, other child favored	.40*	30	Unwise discipline—father	.48**	30
Parents worried over heredity	.40*	30	Child neglected	.47**	33
Unsatisfactory play—mother with child	.38*	31	Parental incompatibility	.43*	29
Mother irritable	.36*	30	Unsatisfactory play—mother with child	.43*	31

* $p < .05$.

** $p < .01$.

*** $p < .001$.

TABLE 4

SIGNIFICANT CORRELATIONS BETWEEN STRUCTURED DEFENSES AND FAMILY ENVIRONMENT

Displacement			Isolation		
Variable	r	N	Variable	r	N
Early family environment		39	Early family environment		39
Conflict over leisure time—mother	.37*		Friendliness of relationship—mother to child	.33*	
Poor health—father	.32*		Family environment at adolescence		
Family environment at adolescence			Strain and conflict in the home	.86**	8
Father restrictive	.51**	29	Straining relationship with child—mother	.53**	31
Unsatisfactory play—mother with child	.37*	31	Child disliked by father—other child favored	.48**	30
Straining relationship—mother with child	.36*	31	Even-temper—father	.44*	27
Projection			Mother demonstrative	.41*	30
Early family environment		39	Agreement over discipline	.40*	30
Help with housework	.35*		Father restrictive	.37*	29
Conflict over leisure time—father	.34*				
Family environment at adolescence					
Straining relationship—mother and child	.41*	31			
Consistent discipline—father	.39*	29			
Mother brutal	.38*	30			
Parents concerned over heredity vs. environment	.38*	30			

* $p < .05$.** $p < .01$.

the family's social status in childhood (Haan, 1964b; Weinstock, 1965) suggests that these variables may play a bigger role in the development of intellectualization than the variables used in the present study.

The results for other obsessive defenses are also difficult to interpret. Isolation appears to be related to a shift from a friendly mother-child relationship in the early family environment to a straining mother-child relationship at adolescence and rejection of the son by the father. Peskin's⁴ findings suggest that subjects in this sample with strong Oedipal attachments to the mother in the early family environment have a straining relationship with the mother and show behavior characteristic of isolation later in their development. The father's dislike of the son at adolescence supports the interpretation that isolation is related to Oedipal conflicts. The father's even temper and the mother's demonstrativeness, as well as the contrast between the childhood and adolescent relationship between the mother and subject, also suggest that the child is faced with striking inconsistencies in the

family which may lead to the compartmentalization of thoughts and feelings characteristic of isolation.

Projection is also related to strain in the mother-child relationship at adolescence and to the mother's brutality. This finding supports Fenichel's (1945) and Swanson's (1961) suggestion that projection is related to realistically perceived rejection and aggression in childhood which leads the child to mistrust others. However, it is not clear from the results why the child should project his impulses on to others.

In general, these results are in agreement with Byrne's (1964) findings that subjects who use obsessional, sensitizing defenses characterize their families as restrictive, rejecting, and as having negative feelings toward one another. The finding that displacement, isolation, and projection are related to strain in the parent-child relationship at adolescence with relatively benign early childhood environments supports Miller and Swanson's (1960) hypothesis that in contrast to denial and repression these are relatively differentiated defenses which depend on the development of more differentiated cognitive abilities.

⁴ Harvey Peskin, personal communication, February 1, 1965.

ties than do denial and repression. This finding contrasts rather sharply with the psychoanalytic notion that projection is a primitive mechanism related to strain and conflict experienced as early as 2 years of age.

Coping Mechanisms

In general the coping mechanisms are less highly correlated with family environment than the defenses. Nine coping mechanisms were rated with sufficient reliability to be compared to the family environment. Of these intellectuality, logical analysis, concentration, and substitution have few correlations significant at $p < .05$. Since these are less than a chance expectancy, and difficult to interpret, the results for these mechanisms will not be reported. These more cognitive mechanisms may be more highly related to the subject's intelligence and to the parents' level of intellectual functioning than to the variables measured in this study.

Expressive coping. Tolerance for ambiguity and regression in the service of the ego have an intercorrelation of .53 and are related to similar family environments. Both mechanisms have little relationship to the early family environment variables, but are related to the father's psychological instability, poor discipline, and conflicts over discipline in ado-

lescence. Tolerance for ambiguity is also slightly related to the father's irritability and explosiveness, in both early childhood and adolescence.

Dauids (1964) and others studying creativity have suggested that tolerance for ambiguity as well as creativity and psychopathology are related to inconsistencies, conflict, and complexity early in development. The results summarized in Table 5 suggest that tolerance of ambiguity and regression in the service of the ego are related to these conflicts and complexities at adolescence and unrelated to variables measuring conflict in the early family environment. These results suggest that the child is able to develop coping resources in a relatively stable early family environment and learns to tolerate conflict and emotional expression in himself and in others through exposure to conflicting, emotionally expressive parents at adolescence. In comparing these mechanisms to regression and doubt, we find these defenses related to similar conflict, and strain in the family at adolescence. However, with the defenses, the father is withdrawing and inactive early in the child's development and the mother is psychologically unstable.

There are few significant correlations between empathy and family environment. The

TABLE 5

SIGNIFICANT CORRELATIONS BETWEEN EXPRESSIVE COPING MECHANISMS AND FAMILY ENVIRONMENT

Tolerance of ambiguity			Regression in the service of the ego		
Variable	r	N	Variable	r	N
Early family environment		39	Family environment at adolescence		
Father irritable	.37*		Father psychologically unstable	.59**	26
Family environment at adolescence			Discipline conflicts	.52**	30
Discipline conflicts	.56**	30	Poor health—father	.51**	31
Father psychologically unstable	.50**	26	Friction over discipline—parents and child	.50**	33
Friction with child over discipline	.47**	33	Parental incompatibility	.50**	29
Poor discipline	.74*	6	Conflict over religion	.47**	29
Conflict over religion	.44*	29	Poor discipline	.82*	6
Father irritable	.41*	27	Empathy		
Parental incompatibility	.41*	29	Early family environment		39
Inconsistent discipline—father	.38*	29	Concern with health—mother	.41**	
Poor health—father	.38*	31	Poor health—mother	.33*	
			Family environment in adolescence		
			Rational discipline—mother	.38*	31
			Mother overprotective	.37*	30

* $p < .05$.

** $p < .01$.

TABLE 6

SIGNIFICANT CORRELATIONS BETWEEN SUPPRESSION, SUBLIMATION, AND FAMILY ENVIRONMENT

Suppression			Sublimation		
Variable	r	N	Variable	r	N
Early family environment		39	Early family environment		39
Concern with health—mother	.44**		Conflict over leisure time—father	.61***	
Mother withdraws in conflict	.34*		Discrepant educational values	.47**	
Agreement over leisure time—mother	.34*		Conflict over leisure time—mother	.43**	
Father withdraws in conflict	.34*		Assertion in conflict—father	.35*	
Family environment at adolescence			Sex adjustment	.33*	
Mother overprotective	.52**	30	Family environment at adolescence		
Mother demonstrative	.49**	30	Child allowed freedom by father	.38*	29
Child allowed freedom by father	.44*	29	Mother sentimental, soft	.38*	30
Mother psychologically stable	.40*	30			
Mother sentimental, soft	.38*	30			

* $p < .05$.** $p < .01$.*** $p < .001$.

mother is concerned over health, uses rational discipline, and is overprotective. These findings suggest that the mother shows the child when he is inconsiderate through rational discipline and encourages him to be concerned with others through her own modeling.

Suppression and sublimation. The suppression rating is a measure of the subject's ability to experience impulses and to also hold them in abeyance until the appropriate moment. The results indicate that the parents allow the child freedom in adolescence yet are protective. The mother is demonstrative, psy-

chologically stable, and soft. These results suggest that the parents encourage and model the expression of socially desirable forms of impulse expression.

In the past sublimation has been used as a general designation for coping. In this study, it has been distinguished from other forms of coping. Like suppression, sublimation is correlated with freedom allowed the child in adolescence. The child is encouraged to develop his own impulse controls. In contrast to suppression, sublimation is highly related to both conflict over leisure time and dis-

TABLE 7

SIGNIFICANT CORRELATIONS BETWEEN DEFENSIVENESS, COPING, AND FAMILY ENVIRONMENT

Defensiveness			Total coping score		
Variable	r	N	Variable	r	N
Early family environment		39	Early family environment		39
Father withdraws in conflict	.42**		Father irritable	.41**	
Even-tempered—father	.38*		Father tense or worrisome	.37*	
Conflict over leisure time—mother	.35*		Agreement over leisure time—father	.32*	
Poor sex adjustment	.35*		Family environment in adolescence		
Mother irritable	.33*		Mother overprotective	.42*	30
Family environment at adolescence			Mother demonstrative	.38*	30
Mother brutal	.58***	30			
Straining relationship—mother and child	.56**	31			
Unsatisfactory play—mother with child	.46**	31			
Strain and conflict in the home	.76*	8			
Father restrictive	.39*	29			
Child neglected	.35*	33			

* $p < .05$.** $p < .01$.*** $p < .001$.

crepant educational values in the home. Despite these disagreements the parents are sexually adjusted. These findings suggest that the parents encourage the child to channel his impulses into constructive activities such as the pursuit of education.

Defensiveness and Coping

Ruebush et al. (1963) found that mothers of defensive males are not supportive and responsive to their sons' needs. These findings are clearly supported by a number of highly significant correlations in the present study between family environment and the defensiveness score (the sum of all the defense ratings). The results summarized in Table 7 indicate that the mothers of highly defensive males are irritable throughout the child's development and are in particular brutal in their relationship to the child at adolescence. In addition, the father is inactive and withdraws in conflict. The father's passivity and relative lack of power in his relationship to his wife seems to be a particularly important determinant of defensiveness in the son. This finding supports Lidz' (1963) contention that an ineffectual father and a cold, unyielding mother will lead to poor sex-role identification and inadequate coping capacities. There are few significant correlations between the degree of coping and the family variables measured. In contrast to the family variables related to defensiveness, these indicate that the father is the more dominant member of the family, and that both parents are emotionally expressive and respect the males in the family.

SUMMARY AND CONCLUSIONS

The results of the present study point up the inadequacy of purely intrapsychic explanations of the origin of ego mechanisms and indicate that the parents' modeling of particular behaviors is an important determinant of the defenses used by the son. Primitive mechanisms like denial, repression, regression, and doubt are closely related to similar behavior in the parent. Other defenses are also related to the child's reaction to the social situation. For example, in displacement the father's restrictiveness is important in deter-

mining the son's displacement of impulse expression.

While the coping mechanisms are also related to the parents' modeling, the relationships between coping and family environment appear to be more complex than for defenses. The results relating family variables to coping mechanisms suggest that the subjects do not merely react to the family environment or model parental behavior, but try actively to master social situations. For example, tolerance of ambiguity and regression in the service of the ego are related to the father's modeling of impulse expression. But these mechanisms are also related to conflict in the home during their adolescent years following a relatively stable early family environment. It appears that subjects exposed to considerable family conflict during adolescence become better able to deal with both external conflict and their own impulses in adulthood.

This finding contrasts with the finding that early family conflict leads to regressive behavior in the subject. Taken together these results suggest that the subject's level of cognitive functioning at the time family difficulties occur plays an important part in determining which ego mechanisms become a permanent part of the subject's character structure. The immature ego seems to react to conflict in the early family environment by rigid imitation of the parents' ways of handling conflict, while the more mature ego learns to confront conflict and impulses and to deal with them in an adaptive way. Erikson (1963) has similarly suggested that less adaptive functioning is related to "overidentifications," resulting from unresolved crises early in development, which hinder the construction of a more flexible and integrated identity at adolescence. Bronson (1959) has demonstrated that stress in the early relationship between boys and their fathers leads to defensive overidentifications or to rejection of masculine attitudes and needs.

The results of this study suggest that future theoretical formulations of the way in which family environment influences character structure should give increased attention to the behavior modeled by the parents and to the demands of the social situation created by

the family. The results also attest to the father's importance in the development of ego mechanisms in males. While psychoanalytic theory has acknowledged the importance of the son's identification with the father after the resolution of the Oedipus complex, theory and research on the early influence of the family have focused on the mother's relationship with the child. The mother seems to play a relatively more important role in adolescence when strain in the mother-child relationship is related to displacement, projection, and isolation.

The results of this study suggest that the way in which the family influences the development of ego mechanisms is quite complex. Many of the findings raise more questions than are answered. While many significant relationships between family environment and ego mechanisms have been found, a large portion of the variance is left unexplained. Murphy et al. (1962) suggest that constitutional factors also play an important role in determining individual differences in the use of ego mechanisms. Further research is needed to clarify the findings of the present study and to account for the way in which other factors including variables derived directly from psychoanalytic theory also contribute to the development of ego mechanisms.

REFERENCES

- BANDURA, A., & WALTERS, R. H. *Social learning and personality development*. New York: Holt, Rinehart & Winston, 1963.
- BRONSON, W. C. Dimensions of ego and infantile identification. *Journal of Personality*, 1959, 27, 532-545.
- BYRNE, D. Childrearing antecedents of repression-sensitization. *Child Development*, 1964, 35, 1033-1039.
- DAVIDS, A. Psychodynamic and sociocultural factors related to intolerance of ambiguity. In R. W. White (Ed.), *The study of lives*. Englewood Cliffs, N. J.: Prentice-Hall, 1964. Pp. 160-177.
- ERIKSON, E. H. *Childhood and society*. New York: Norton, 1963.
- FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.
- HAAN, N. Proposed model of ego functioning: Coping and defense mechanisms in relationship to IQ change. *Psychological Monographs*, 1963, 77(8, Whole No. 571).
- HAAN, N. An investigation of the relationship of Rorschach scores, patterns, and behavior to coping and defense mechanism. *Journal of Projective Techniques and Personality Assessment*, 1964, 28(4), 429-441. (a)
- HAAN, N. The relationship of ego functioning and intelligence to social status and social mobility. *Journal of Abnormal and Social Psychology*, 1964, 69, 594-605. (b)
- ETHERINGTON, E. M., & BRACKBILL, Y. Etiology and covariation of obstinacy, orderliness, and parsimony in young children. *Child Development*, 1963, 34, 919-943.
- KROEBER, T. C. The coping functions of the ego mechanisms. In R. W. White (Ed.), *The study of lives*. New York: Atherton Press, 1963. Pp. 178-199.
- LIDZ, T. *The family and human adaptation*. New York: International Universities Press, 1963.
- MACFARLANE, J. W. Methodology of data collection and organization, studies in child guidance. *Monographs of the Society for Research in Child Development*, 1938, 3, No. 6.
- MILLER, D. R., & SWANSON, G. E. *Inner conflict and defense*. New York: Holt, 1960.
- MURPHY, L. B., ET AL. *The widening world of childhood: Paths toward mastery*. New York: Basic Books, 1962.
- RUEBUSH, B. K., BYRUM, M., & FARNHAM, L. J. Problem solving as a function of children's defensiveness and parental behavior. *Journal of Abnormal and Social Psychology*, 1963, 67, 355-362.
- SWANSON, G. E. Determinants of the individual's defenses against inner conflict: Review and reformulation. In J. C. Glidewell (Ed.), *Parental attitudes and child behavior*. Springfield, Ill.: Charles C Thomas, 1961. Pp. 5-41.
- WEINSTOCK, A. R. Family environment and ego functioning: A longitudinal study of the family antecedents of defense and coping mechanisms. Unpublished doctoral dissertation, University of California, Berkeley, 1965.

(Received March 14, 1966)

AUTOKINETIC NORMS: AN EXPERIMENTAL ANALYSIS¹

WILLIAM F. STONE²

University of Florida

Male Ss formed autokinetic norms under 1 of 9 conditions of practice (20, 40, or 80 trials) and group size (1, 2, or 4 persons). In a subsequent session, the strength of Ss' norm was tested by means of a verbal conditioning procedure. It was found that: (a) Mean estimates (norms) vary inversely with group size ($p < .01$); (b) the conditioning procedure was effective in changing Ss' estimates ($p < .001$); (c) group size was related to the strength of Ss' norms—Ss who had formed norms in the 1-, 2-, and 4-person groups were successively more likely to return to their original norms ($p < .01$); (d) practice had no effect. The term "group reinforcement" seems appropriate to describe the effects of group judging, but the mechanism by which mutual reinforcement occurs is unclear.

Although there has been controversy (Gibson, 1953) concerning the generality of his results, Sherif's (1935) study of norm formation in the autokinetic situation has been a continuing stimulus to research in social cognition. A peculiar advantage is that the judgment of autokinetic movement is a task new to most subjects. Thus, it is possible to observe the formation of novel opinions in the laboratory. Elaborations on the use of the autokinetic situation in the study of social judgment have been made by Sherif and Hovland (1961), Jacobs and Campbell (1961), and others.

There is evidence that social ("group") reinforcement is operative in the autokinetic situation. Sherif (1935), for instance, demonstrated that perceptual norms are more stable when formed with another person or

persons. Subsequent investigators have found that autokinetic judgments are more likely to remain centered on the same value if the subject is told that his judgments are accurate (Harvey & Rutherford, 1958; Kelman, 1950; Mausner, 1954). These findings accord well with the assertion that reward serves to fix perceptual responses (Solley & Murphy, 1960). Although Solley and Murphy also suggest that perceptual responses are self-reinforcing, there seems to be little evidence concerning the effects of practice (i.e., the number of self-reinforcements) in setting perceptual norms.

The present study examined the effects of practice and group size on perceptual norm formation. It was hypothesized that:

1. When autokinetic norms are learned in the absence of external reward, there is an increment of habit strength associated with each response, because "the achievement of percepts is reinforcing to a perceptual act [Solley & Murphy, 1960, p. 82]."

2. With the addition of other observers, habit increments will be larger, owing to "group reinforcement." That is, hearing other persons make responses similar to his own will be reinforcing and hence will increase the strength of the subject's response. As the habit strength of the responses near the norm increases, it is expected that the norm will become stronger. The strength of norms formed under the various conditions was ex-

¹ Based upon a dissertation submitted to the graduate school, University of Florida. Partial support was afforded by a Veterans Administration Traineeship held by the author, and by a grant from the Research Committee, Veterans Administration Hospital, Coral Gables, Florida. Marvin E. Shaw, who directed the dissertation, also read a draft of this article and made many helpful suggestions. Appreciation is expressed to J. McDavid and other faculty at the University of Miami for subjects and use of facilities. W. B. Webb, H. D. Kimmel, and J. Sandler gave valuable assistance in the design and execution of the study. This article was prepared while the author was at Lafayette College.

² Now at the University of Maine.

pected to be reflected in the subject's behavior in an operant conditioning situation (Kanfer, 1954; Spivak & Papajohn, 1957) imposed on the day following the formation of his norm.

METHOD

Subjects

The subjects were 150 male undergraduates enrolled in psychology classes at the University of Miami. They received course credit for participation.

Apparatus

The autokinetic light was mounted at the rear of a light-tight plywood box, 4 feet high \times 5½ feet wide. Four cloth-hooded viewing ports were provided on the four panels (which were set at angles of 135 degrees to each other) which comprised the front of the box. The spot of light was produced by means of a ¼-inch diameter lucite rod, which protruded ½ inch upward from a small metal box. A ½-watt neon bulb in the metal box was operated at 132 volts to illuminate the rod, which presented a ½-inch square target 2½ feet from each subject. Timing relays provided 7-second exposures alternating with 13-second dark periods.²

The experimental room was dimly illuminated by two red 25-watt bulbs, and an air-conditioner to the rear of the apparatus ran constantly.

Design of the Experiment

The experiment was conducted in two periods, on successive days. On the first day, the subjects formed norms under the various experimental conditions (the norm session). On the second day, subjects were run individually in the conditioning and extinction sessions.

The main experiment consisted of nine conditions, representing all possible combinations of three group sizes (one, two, and four) and three numbers of practice trials during norm formation (20, 40, and 80). That is, during the norm session on Day 1, subjects made judgments of autokinetic movement either alone, with another person, or with three other persons, and made either 20, 40, or 80 judgments. On Day 2, all subjects were treated alike. They were given up to 40 conditioning trials during which 10 reinforcements for high responses were given, followed by 40 extinction trials.

In addition to the experimental groups, two control groups were run. Control groups made 40 judgments alone in the norm session. Control Group A also formed norms alone, but received random reinforcements from the experimenter, who said "right" following 50% of the judgments. This group was treated the same as the experimental groups during

the conditioning and extinction sessions. Control Group B received no reinforcements during the norm session. During the conditioning session, subjects in this group received 10 noncontingent reinforcements, determined by yoking each subject with a subject in the comparable experimental condition, 40-1 (group size, 1, 40 judgments). Extinction was the same as in other groups.

Procedure

The subjects were recruited for the "night vision experiment." It was announced, in the call for volunteers, that the experiment involved "making judgments of the movement of a small light in a dark field."

Day 1 (norm session). The subjects entered the laboratory, chose seats at one of the portholes, and during the 5-minute dark adaptation period were told:

This is an experiment in visual discrimination. After I tell you what to do, and you take your places at the apparatus, I will turn on a small light inside. After a short while, the light will start to move. Then it will go off. After the light is off I want you to tell me the distance the light has moved. Make your estimates in full inches or half inches. For example, say "5½ inches," or "7½ inches," or "10 inches," and so on. (Give your estimates one at a time, beginning with the person at the left.) The light will move no less than 1 inch and no more than 24 inches. Report only the distance that the light moves. Do not report the direction of movement [adapted from Kanfer, 1954].

After a practice trial the stimulus was presented periodically, and the subjects reported their estimates in turn during the 13-second intertrial interval. A 1-minute rest period followed every twentieth exposure. Upon completion of the trials, individual appointments were made for the following day.

Based upon the distribution of his last 20 estimates, the high response class for each subject was defined as any response above his median class.

Day 2 (conditioning and extinction sessions). The subject entered the laboratory alone, dark adapted for 5 minutes, and was told: "I want you to make estimates as you did yesterday. At the start, I'll say 'right' when you guess within ½ inch of the correct distance. Otherwise, I'll say nothing . . ."

The first 10 high responses emitted by the subject were reinforced, on a continuous schedule. The session was terminated after the tenth reinforcement, or the fortieth trial, whichever came first. Subjects who did not make 10 high responses within 40 stimulus presentations were discarded. After a 2-minute rest period the extinction session began. No additional comments, instructions, or verbal reinforcements were given. Forty extinction trials were run (2 blocks of 20 estimates each, separated by a 1-minute rest period). Following the fortieth trial, a brief interview

² A sketch and more detailed instructions for construction of the apparatus may be obtained from the author.

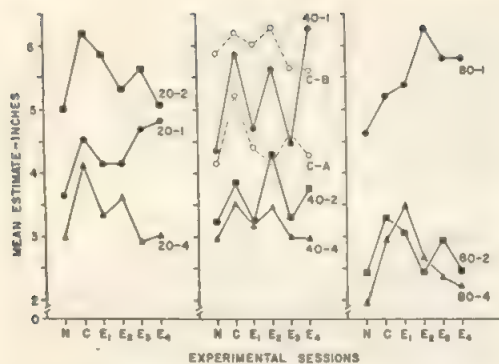


FIG. 1. Conditioning and extinction of autokinetic judgments as functions of the learning conditions. (See text for explanation of symbols.)

was held with the subject, and two paper-and-pencil questionnaires were administered.

The conditioning and extinction procedure for Control Group A was the same as for the experimental groups. Group B was also treated the same, except that the reinforcements for Group B subjects were not contingent on their responses. Each subject in Group B was "yoked" with one subject in Group 40-1, in that the B subject was given reinforcements on the same trials as had his yokedmate.

RESULTS

A total of 138 subjects returned for the second day's sessions out of 150 who participated in the norm-formation procedure. Of these, 9 failed to meet the conditioning criterion (4 from Group 40-1, 1 each from 5 other groups). Seven additional subjects were over the preset age limit of 30 years and were dropped, and data from 12 subjects were randomly discarded to maintain equal cell frequencies in each of the 11 groups.

The session-by-session estimates for the 10 subjects in each of the experimental groups are shown in Figure 1. Norm-session (N) scores are means of the subject's last 20 estimates, C scores are means of all of his estimates in the conditioning session, and E scores are means of successive 10-trial blocks of estimates in the extinction session.

Norms. Although the norms for individual subjects are the basis for subsequent analyses, it should be noted that those who formed norms together did form *group* norms, as expected. For example, the mean difference between N scores of members of the two-man

groups was .77 inch, as compared with 3.50 inches for random pairs formed from subjects who judged alone.

It is apparent (Figure 1) that there were large differences among the various groups in mean N scores, and that these between-group differences in norms were perpetuated throughout the experimental sequence. Means and SDs of N scores for the experimental groups are shown in Table 1. Initial differences in norms were tested by the between-subject terms of a mixed analysis of variance of the N and C data. A significant group-size F ratio ($F_{2,81} = 5.81$, $p < .01$) demonstrated that mean estimates (norms) of individual subjects are inversely related to size of group in which they formed their norms. Changes over sessions were not, however, correlated with N scores. Therefore the absolute differences between groups were ignored in subsequent analyses, and attention was focused on within-subject changes over sessions.

The SDs of the experimental group means which are reported in Table 1 also reflect between-group differences in variability. This heterogeneity tended to be perpetuated through all experimental periods, and to allow for it apparent significance levels were doubled in making interpretations (Lindquist, 1953, p. 83). F ratios having a probability of .025 or less were considered statistically significant.

TABLE 1

MEAN ESTIMATES (IN INCHES) FOR NORM SESSION
(BASED ON MEANS OF SUBJECTS'
LAST 20 JUDGMENTS)

Practice trials	Group size		
	1	2	4
20			
<i>M</i>	3.65	5.02	2.99
<i>SD</i>	2.95	1.72	1.30
40			
<i>M</i>	4.37	3.25	2.97
<i>SD</i>	1.60	1.26	0.69
80			
<i>M</i>	4.65	2.43	1.95
<i>SD</i>	2.73	1.18	0.72

Note.— $N = 10$ subjects per group.

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE OF EXTINCTION DATA (FOUR 10-TRIAL BLOCKS)

Source	df	MS	F
Between Ss	89		
Practice (P)	2	12.12	0.71
Group size (G)	2	141.37	8.36***
P \times G	4	46.52	2.75
Error (b)	81	16.91	
Within Ss	270		
Trend (T)	3	1.44	1.35
P \times T	6	2.86	2.67*
G \times T	6	3.35	3.13**
P \times G \times T	12	0.44	0.41
Error (w)	243	1.07	
Total	359		

* $p < .025$.** $p < .01$.*** $p < .001$.

Conditioning. The within-subject analyses of the N-C data showed quite conclusively that: (a) The conditioning procedure was effective in changing the subjects' mean estimates (trend $F_{1, 81} = 57.76, p < .001$); and (b) there were no differential changes among groups from N to C (no trend interactions). Thus there was no evidence of between-group differences in conditionability. (This finding is not affected by the initial differences in norms, since the correlations between individual norms and changes from N to C do not differ from zero.)

Extinction. The differential changes in slope of the extinction curves of the experimental groups were reflected in the trend interactions of the analysis of variance presented in Table 2. In general, subjects who form norms alone show greater resistance to extinction than do those who form norms in groups of two or four, as shown by the significant Group Size \times Trend interaction. In fact, analysis of trend effects for the three group sizes taken separately shows that subjects who form their norms alone show an increasing trend in their estimates over the extinction session ($F_{3, 243} = 4.37, p < .01$). Subjects from the two-man groups show a downward, though not significant, trend ($F_{3, 243} = .47$). Those who form norms in the four-man groups show near significant decreases ($F_{3, 243} = 2.78, p < .05$), a tendency to return to their norms.

The Trend \times Practice interaction shown in Table 2 seems to be accounted for entirely by the periodic vacillation in estimates of subjects in the 40-trial groups, as shown by the extinction trend for the combined 40-trial groups, tested against error (w) ($F_{3, 243} = 5.46, p < .005$). Neither the 20-trial nor the 80-trial groups showed significant trends over the extinction session.

Control groups. Means for the two control groups, A and B, are plotted in the center panel of Figure 1. Comparison of these groups with 40-1 was made by an analysis over all six blocks. This analysis shows that: (a) The absolute differences between groups were not significant ($F < 1$); and (b) the trend for all three groups combined approached significance ($F_{5, 135} = 2.25, p < .10$). When the trend for each group was tested against error (w), neither Group A nor Group B showed F ratios greater than 1. The trend for 40-1 was significant ($F_{5, 135} = 3.81, p < .005$). Neither Group A, in which the subjects' norm-session estimates were randomly reinforced by the experimenter, nor Group B (noncontingent reinforcement in the conditioning session) showed changes from their own norms. These analyses, plus the striking resemblance of the control-group curves to those for 40-2 and 40-4, suggest that: (a) Reinforcement by the experimenter during norm formation produces resistance to change similar to that produced by judging with others; and (b) changes in estimates which occur during conditioning are produced by the contingency of the reinforcements on high responses, and do not result from "encouragement" or the like from the experimenter's vocalizations.

DISCUSSION

Norm formation. Subjects in the group norm-formation conditions (in addition to the oft-noted convergence of estimates) showed a decided tendency to report less movement than do subjects judging alone. This phenomenon is of interest, since it suggests a tendency for the observers to be influenced in the direction of "reality" (the light actually does not move). Subjects who experience relatively less movement seem to have more

influence in determining the group norm. A somewhat similar finding was reported by Jacobs and Campbell (1961), who found that an arbitrary norm of 15 inches gradually decayed to about 3 inches as naive subjects replaced those who had been indoctrinated by confederates. These authors postulate the existence of a natural norm for autokinetic movement, averaging about 3 inches in their experimental setting (8-foot viewing distance, 5-second exposure).

The shift toward zero in groups judging autokinetic movement, together with the findings concerning shifts of member opinion toward riskier positions (Wallach, Kogan, & Bem, 1962), brings into question certain assumptions about opinion convergence in groups. French (1956), for example, suggests that the most probable value for the "final common opinion" reached in a group will be the arithmetic mean of the initially divergent opinions. It seems more probable that in a given situation opinions at certain positions on the continuum carry more weight than others. In general, when initial opinions are well distributed, there will be convergence, but with some polarization toward one or the other extreme (in the autokinetic situation, toward zero).

Practice and reinforcement. The present study confirms previous findings (Harvey & Rutherford, 1958; Kelman, 1950) concerning the strength of autokinetic norms as a function of prior reinforcement. Subjects who form norms in groups retain them longer than do those who judge alone initially. Maximum strength appears to be attained in the four-man groups, whose members' norms are somewhat stronger than those of subjects in the two-man groups. The major behavioral differences in extinction, however, are in the contrast between subjects who form norms alone and those who had at least one other person judging with them.

Surprisingly, amount of practice has no demonstrable relationship to norm strength. This finding is of interest since two considerations lead us to predict increased strength of autokinetic norms with practice. First, the hypothesized "self-reinforcement" of perceptual responses is expected to contribute to

habit strength in proportion to the number of responses. Second, within-subject variability of autokinetic estimates decreases over trials, and subjects who gave more variable estimates (i.e., who had fewer learning trials) should be more influenceable (Bovard, 1948).

Group reinforcement. The present study has demonstrated that norm responses made in the presence of others are strengthened over those made alone. Further, it was shown that the "judging-with-others" effect can be duplicated by delivery of verbal reinforcements to solitary respondents. These findings appear to confirm the notion that autokinetic norms are formed through the reinforcement of "reporting responses" (Schoenfeld & Cumming, 1963). However, the manner in which the subjects reinforce one another's responses is not altogether clear, since the order of response and reinforcement is not always that specified by the instrumental learning model. While the existence of an empirically defined group reinforcement has been demonstrated, its mediating mechanism remains to be explicated.

REFERENCES

- BOVARD, E. W., JR. Social norms and the individual. *Journal of Abnormal and Social Psychology*, 1948, 43, 62-69.
- FRENCH, J. R. P., JR. A formal theory of social power. *Psychological Review*, 1956, 63, 181-194.
- GIBSON, J. J. Social perception and the psychology of perceptual learning. In M. Sherif & M. O. Wilson (Eds.), *Group relations at the crossroads*. New York: Harper, 1953. Pp. 120-138.
- HARVEY, O. J., & RUTHERFORD, J. Gradual and absolute approaches to attitude change. *Sociometry*, 1958, 21, 61-68.
- JACOBS, R. D., & CAMPBELL, D. T. The perpetuation of an arbitrary tradition through several generations of a laboratory microculture. *Journal of Abnormal and Social Psychology*, 1961, 62, 649-658.
- KANFER, F. H. The effect of partial reinforcement on acquisition and extinction of a class of verbal responses. *Journal of Experimental Psychology*, 1954, 48, 424-432.
- KELMAN, H. C. Effects of success and failure on "suggestibility" in the autokinetic situation. *Journal of Abnormal and Social Psychology*, 1950, 45, 267-285.
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
- MAUSNER, B. The effect of prior reinforcement on the interaction of observer pairs. *Journal of Abnormal and Social Psychology*, 1954, 49, 65-68.

SCHOENFELD, W. N., & CUMMING, W. W. Behavior and perception. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 5. New York: McGraw-Hill, 1968. Pp. 213-252.

SHERIF, M. A study of some social factors in perception. *Archives of Psychology*, 1935, No. 187.

SHERIF, M., & HOSLAND, C. I. *Social judgment*. New Haven: Yale University Press, 1961.

SOLLEY, C. M., & MURPHY, G. *Development of the perceptual world*. New York: Basic Books, 1960.

SPIVAK, M., & PAPAJOHNN, J. The effect of schedule of reinforcement on operant conditioning of a verbal response in the autokinetic situation. *Journal of Abnormal and Social Psychology*, 1957, 54, 213-217.

WALLACH, M. A., KOGAN, N., & BEM, D. J. Group influence on individual risk taking. *Journal of Abnormal and Social Psychology*, 1962, 65, 75-86.

(Received March 22, 1966)

Manuscripts Accepted for Publication in the *Journal of Personality and Social Psychology*

Situational Determinants of Affiliative Preference under Stress: Robert L. Helmreich and Barry E. Collins*: Department of Psychology, Yale University, New Haven, Connecticut.

Cognitive Control and Cognitive Dissonance: David L. Wolitzky*: Research Center for Mental Health, New York University, 4 Washington Place, New York, New York 10003.

Variations in Manifest Anxiety due to Sex, Age, and Education: Vasso Vassiliou, James G. Georgas,* and George Vassiliou: Athenian Institute of Anthropos, 8, Dem. Soutsou Street, Athens 602, Greece.

Development of Contractual Norms in a Dyad: Peter Murdoch*: Department of Psychology, University of Iowa, Iowa City, Iowa 52240

A Study in Time Orientation: The Relation between Memory for Past Experience and Orientation to the Future: Judith March Goldrich*: 96 Fifth Avenue, New York 11, New York.

Punishment and Aggression in Fantasy Responses of Boys with Antisocial Character Traits: Judith B. Schaefer*: Newton-Baker Project, 64 Washington Park, Newtonville, Massachusetts.

Fear Reduction Following Observation of a Model: James H. Geer* and Alan Turteltaub: Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania 19104.

The Anal Character: A Rebel in the Dissonance Family: Frances V. Bishop*: 59 West 12th Street, New York, New York 10011.

Anxiety and Repression as Predictors of the Use of Incidental Cues in Problem Solving: Gerald A. Mendelsohn* and Barbara B. Griswold: Department of Psychology, University of California, Berkeley, California 94720.

Personal Religious Orientation and Prejudice: Gordon W. Allport* and J. Michael Ross: Department of Social Relations, Harvard University, Cambridge, Massachusetts 02138.

Bargaining, Expectations, and the Preference for Equality over Equity: William R. Morgan and Jack Sawyer*: Department of Sociology, University of Chicago, 1126 East 59th Street, Chicago, Illinois 60637.

Some Social Determinants of Self-Monitoring Reinforcement Systems: Albert Bandura,* Joan E. Grusec, and Frances L. Menlove: Department of Psychology, Stanford University, Stanford, California.

Athletic Participation and Perceptual Augmentation and Reduction: E. Dean Ryan*: Department of Physical Education, University of California, Davis, California 95616.

Effects of Pretraining on Self-Evaluation and Self-Reinforcement: Frederick H. Kanfer* and Pryse H. Duerfeldt: Department of Medical Psychology, University of Oregon Medical School, 3181 S. W. Sam Jackson Park Road, Portland, Oregon 97201.

Attitudinal Effects of Communicator Attractiveness When One Chooses to Listen: Russell A. Jones and Jack W. Brehm*: Department of Psychology, Duke University, Durham, North Carolina.

Changes in Outlook on the Future between Childhood and Adolescence: Stephen L. Klineberg*: Department of Sociology, Princeton University, Princeton, New Jersey 08540.

Effects of Fear and Instructions on How to Cope with Danger: Howard Leventhal,* Jean C. Watts, and Francia Pagano: Department of Psychology, Yale University, New Haven, Connecticut.

Models and Helping: James H. Bryan* and Mary Ann Test: Center for Psychological Studies, Educational Testing Service, Princeton, New Jersey 08540.

* Asterisk indicates author for whom address is supplied.

ATTRACTION AND SIMILARITY OF PERSONALITY CHARACTERISTICS¹

DONN BYRNE, WILLIAM GRIFFITT, AND DANIEL STEFANIAK

University of Texas

It was hypothesized that attraction toward another individual is a positive linear function of the proportion of his personality characteristics which are similar to those of S. In the 1st experiment, 151 Ss examined the responses of a stranger to the Repression-Sensitization (R-S) Scale. The stranger responded as S did on .20, .50, or .80 of the items. Analysis of variance indicated that attraction was affected by proportion of similar responses ($p < .001$) and by repression-sensitization ($p < .01$). In a 2nd experiment, employing 149 Ss, attitude similarity influenced attraction ($p < .001$), but repression-sensitization did not. In the 1st experiment, subject-stranger discrepancy in R-S Scale scores was found to influence attraction ($p < .001$); represser-sensitizer differences were thus artifactual. The relationship between personality similarity and attraction is entirely consistent with the findings of attitude-similarity studies.

Among the proposed determinants of interpersonal attraction between any two individuals is the extent to which they are similar in personality, with personality operationally defined in terms of scores on one or more measures of individual differences. In the present paper, an attempt is made to integrate personality similarity into the more general framework of a reinforcement theory of attraction.

In a series of investigations in which attitudinal similarity between each subject and a stranger has been experimentally manipulated, a linear function between proportion of similar attitudes and attraction has repeatedly been found (e.g., Byrne & Clore, 1966; Byrne & Griffitt, 1966; Byrne & Nelson, 1965). Attitude statements are conceptualized as positive and negative reinforcements in that they provide evidence of one's effectiveness in perceiving and interpreting his stimulus world (Byrne, Nelson, & Reeves, 1966; Golightly & Byrne, 1964; McDonald, 1962; Nelson, 1965). The empirically derived linear function led to the postulation of the *law of attraction*: attraction toward X is a positive linear function of the proportion of positive reinforcements received from X: $(A_x = mPR_x + k)$.

¹ This research was supported in part by Research Grant MH-11178-01 from the National Institute of Mental Health, United States Public Health Service. The authors wish to thank Dennis Campbell, Rudy Garcia, and Kay Wong for their assistance in this research.

The applicability of this formulation to personality similarity is theoretically contingent on the positive or negative reinforcing properties of information about another individual's personality characteristics.

Festinger (1954, p. 118), in his theory of social comparison processes, hypothesized that "To the extent that objective, non-social means are not available, people evaluate their opinions and abilities by comparison respectively with the opinions and abilities of others." He goes on to suggest that "A person will be less attracted to situations where others are very divergent from him than to situations where others are close to him for both abilities and opinions [p. 123]." The hypothesized positive relationship between similarity and attraction has been supported not only with respect to opinions (Byrne & Nelson, 1965) and abilities (Zander & Havelin, 1960), but also with emotional states (Zimbardo & Formica, 1963), self-description (Griffitt, 1966), evaluation of the subject's performance (Deutsch & Solomon, 1959), economic status (Byrne, Clore, & Worchel, 1966), and even with amount of paint consumption (Brock, 1965). Thus, the effects of Festinger's (1950, 1954) proposed self-evaluative drive and other related motivational constructs (Byrne, Nelson, & Reeves, 1966) seem to extend beyond opinions and abilities. It is tentatively suggested that, with other variables held constant, the behavior of an-

other individual is positively reinforcing to the extent that it is similar to one's own behavior. Behavioral similarity to self, whether involving attitudes or values or abilities or emotional responses or tastes or adjustive responses or worries or need hierarchies or whatever, provides evidence that one is functioning in a logical and meaningful manner (Byrne, 1961; Cohen, Stotland, & Wolfe, 1955), and it makes one's interpersonal environment more predictable and understandable (Brim & Hoff, 1957; Pervin, 1963). It might be noted that in a sufficiently complex situation, the activation of quite different motive states could probably overshadow such cognitive concerns.

On the basis of this general reasoning, the law of attraction should be applicable to personality similarity⁴ at least within the delimited laboratory setting. It is hypothesized, then, that attraction toward another individual is a positive linear function of the proportion of that individual's personality characteristics which are similar to the characteristics of the subject. Before proceeding to an experimental test of this hypothesis, it is necessary to examine previous research on personality similarity and attraction and to indicate the methodological inadequacies of such research for testing the basic hypothesis.

PREVIOUS RESEARCH ON PERSONALITY SIMILARITY

There has been a considerable degree of theoretical controversy concerning the nature of the relationship between personality similarity and attraction. It has been proposed that attraction is facilitated by either similarity, complementarity, or both (e.g., Levinger, 1964). The theoretical differences remain unresolved because the research findings have been sufficiently inconsistent as to provide support for the similarity hypothesis (Banta & Hetherington, 1963; Izard, 1960a, 1960b; Maisonneuve, 1954; Mehlman, 1962; Miller, Campbell, Twedt, & O'Connell, 1966; Murstein, 1961), for the complementarity hypothesis (Cohen, 1956; Kerckhoff & Davis, 1962; Rychlak, 1965; Winch, Ktsanes, & Ktsanes, 1955), and for some combination of the two (Becker, 1964; Secord & Backman, 1964). There are additional findings which

give only partial support to the similarity hypothesis in that similarity is positively associated with attraction only under limited conditions, or only in specific groups, or with respect to only a few variables (Bonney, 1946; Bowerman & Day, 1956; Izard, 1963; Katz, Glucksberg, & Krauss, 1960; Rosenfeld & Jackson, 1965; Van Dyne, 1940). In addition, a number of investigators have simply found no relationship between personality similarity and attraction (Corsini, 1956; Gordon, 1957; Hoffman, 1958; Hoffman & Maier, 1966; Katz, Cohen, & Castiglione, 1963; Pintner, Forlano, & Freedman, 1937; Reilly, Commins, & Stefci, 1960; Thorpe, 1955).

One reason for empirical inconsistency is the peculiar penchant of personality and social psychologists for methodological creativity such that almost every investigation represents an exploration in procedural novelty. That is, different investigators employ different independent and dependent variables in situations of varying complexity in which a seemingly limitless array of parameters is operative. Over and above this familiar problem, however, are basic design flaws which make it literally impossible to determine the effect of personality similarity on attraction. Two general designs have been utilized. In one approach, existing "real life" attraction pairs (e.g., friends, fiancés, spouses) are selected and assessed with respect to one or more personality variables. Then, the scores of the pairs are cofrelated. These correlations are often compared with similar correlations for random pairs from the same population or with pairs of mutually antagonistic or mutually indifferent subjects. In the second approach, the personality measure or measures are obtained, and then previously unacquainted subjects are selected on the basis of test scores and placed in a situation requiring some degree of interaction. Thus, similar and dissimilar pairs or groups are created, and their interpersonal responses are assessed following the interaction. Two difficulties are inherent in either approach.

First, attraction responses are known to be multidetermined. For example, attraction varies as a function of such determinants as propinquity (e.g., Byrne & Buehler, 1955; Festinger, Schachter, & Back, 1950), the re-

enforcing properties of the situation (e.g., Lott & Lott, 1960), attitude similarity-dissimilarity (e.g., Byrne & Nelson, 1965; Newcomb, 1956), perceived similarity (e.g., Levinger & Breedlove, 1966; Lundy, 1958), the temporal length of the relationship (e.g., Morton, 1960), characteristics like boastfulness and self-depreciation (Pepitone, 1964), the respective status of each individual (e.g., Jones, 1964), etc. In the investigations of personality similarity, all such variables plus an unknown number of as yet unidentified variables are potentially operative, but are not consistently controlled. Further, personality similarity itself is defined in terms of similarity on one or at best a small subgroup of personality variables so that similarity along all other personality dimensions is also not controlled. The situation, then, is one in which the effect of a very limited number of independent variables on attraction is determined in a context where 1— n uncontrolled independent variables are operating. Only if the variable under investigation were of sufficient strength to override all other independent variables or if a sufficient number of the other independent variables happened to covary with it or if the other variables were accidentally controlled through randomization would the hypothesized relationship be observed. Since the obtained findings lack consistency, these special circumstances do not appear to occur regularly. It should be noted that the research approach just described is a perfectly suitable one for answering the limited question concerning the role of similarity along a given personality dimension in a specific uncontrolled realistic situation. This approach is inappropriate, however, for answering the prototypic question concerning the effect of personality similarity on attraction. The solution lies either in employing a multivariate methodology in the uncontrolled situation or in moving to a traditional experimental situation. The latter choice means, of course, a narrow specification in terms of the variables which are employed, the parameters which are investigated, and the situations to which the findings may reasonably be generalized. Strangely enough, the history of science suggests that the apparent artificiality and narrowness of the laboratory provide the

most useful means whereby the complexities of the nonlaboratory world may be clarified.

Even if extraneous stimulus variables are brought under experimental control (e.g., Altrocchi, 1959), a second difficulty remains. The relationship between scores on a personality measure and behavior in an interpersonal situation is crucial. Presumably, personality variables influence attraction only in terms of the effect of such variables on behavior in a given situation; it is the behavioral stimuli to which the subjects are responding and not to the hypothesized personality dimension. But, the relationship between scores on a personality test and behavior in various everyday life encounters or behavior in a controlled laboratory situation is for the most part unknown. It seems a safe bet that a 1:1 relationship does not hold. Thus, one is in the position of defining personality similarity in terms of Behavior A (responses to the test) and determining its effect on attraction in a situation involving Behavior B (responses in the interaction) with no knowledge of the relationship, if any, between Behaviors A and B. In other words, it is necessary to identify the stimulus to which subjects are expected to respond. One must move from an interest in similarity of *personality* to an interest in similarity of *specific personality characteristics*.

It follows from this analysis that an appropriate test of the proposed relationship between personality similarity and attraction will require an experimental design in which (a) there is control of the stimulus determinants other than the independent variable, and (b) there is a known relationship between the personality-relevant behavior of the subject and the personality-relevant behavior of the person to whom he responds. It also follows that any single such investigation cannot be interpreted as indicating the relationship between any other personality characteristics and attraction or the relationship which might be found in quite different situations in which different variables operate. The first experiment to be reported represents one such solution to the design problem and an initial step in establishing a more general relationship between "personality similarity" and attraction.

EXPERIMENT I: SIMILARITY IN DEFENSIVE BEHAVIOR AND ATTRACTION

The personality dimension is that of repression-sensitization, which describes a behavioral continuum involving characteristic responses to anxiety-evoking stimuli. At one extreme are behaviors such as denial, avoidance, and repression, while at the other extreme are behaviors such as intellectualization, approach, and sensitization. Much of the research on this personality dimension has been reviewed elsewhere (Byrne, 1964).

To investigate personality similarity with respect to repression-sensitization or any other dimension, either of two behavioral samples may be employed. First, one might determine the characteristic responses of a subject in a series of quite specific situations and then expose him to the behavior of a stranger in those same situations. Second, the subject's behavior might consist of his specific test responses which were utilized to measure the dimension, and the observed behavior of the stranger would consist of his responses to the same instrument. The latter approach was utilized in the present investigation.

Method

The personality dimension was measured by means of the Repression-Sensitization (R-S) Scale (Byrne, Barry, & Nelson, 1963) which consists of 182 MMPI items of which 127 are scorable while 55 serve as buffers. A special version of the test was constructed for this experiment. In order to limit the stimulus information about the stranger to personality-relevant material, only the 127 cross-validated scorable items were used. In order to facilitate the association between item content and the stranger's responses, answers were made directly on the test booklet with checks in T and F boxes to the left of each item rather than on an IBM answer sheet. The sampled behavioral domain of both subject and stranger thus consisted of 127 responses to the same set of test items. Examples are: "Once in a while I think of things too bad to talk about"; "I have very few quarrels with members of my family"; "I am happy most of the time"; "At times I feel like picking a fist fight with someone." For each item, the responses of the subject and the stranger could either be identical or opposite.

The modified R-S Scale was administered to approximately 450 male and female students enrolled in a section of the introductory psychology course at the University of Texas. On the basis of scores obtained on the test, three groups of subjects were selected: 48 sensitizers (scores of 59-110), 54 neutrals

(scores of 38-45), and 49 repressors (scores of 2-25). There were 151 subjects in all.

Several weeks after the initial test administration, the subjects were seen in small groups for the experiment itself. As in previous investigations, they were told that the experiment dealt with the accuracy of interpersonal judgment. They received a test, purportedly that of another student (name removed) of their same sex. The "stranger's" test was actually filled out by the experimenter. The task was to read through each of the 127 items and study the responses made by the other person. Afterward, subjects were asked to evaluate on the Interpersonal Judgment Scale (six 7-point rating scales) the stranger's intelligence, knowledge of current events, morality, and adjustment and were asked to indicate how much they would like the person and how much they would enjoy working with him. Responses to the latter two items are summed to yield the measure of attraction which ranges from 2 to 14 with a split-half reliability of .85 (Byrne & Nelson, 1965).

Within each of the three repression-sensitization groups, subjects were assigned to one of three experimental conditions, consisting of three levels of subject-stranger similarity in responding to the R-S Scale. In the .20 condition, subjects received a test on which the stranger responded exactly as the subject did to 25 items and exactly the opposite to 102 items. In the .50 condition, there were 64 similar and 63 dissimilar responses. In the .80 condition, there were 102 similar responses and 25 dissimilar ones. Within each condition, a different random pattern of specific items of similarity and dissimilarity was devised for each subject; for example, in the .20 condition each subject responded to a stranger like himself on 25 items, but the actual items comprising the 25 were different for each subject.

Results

The means and standard deviations of the attraction responses for the three repression-sensitization groups in the three experimental conditions are shown in Table 1. The results of a 3×3 factorial analysis of variance, corrected for disproportionality (Wert, Neidt, & Ahmann, 1954), are shown in Table 2. As hypothesized, there is a highly significant effect attributable to differences in the proportion of similar responses to the personality items. In a previous investigation in which subjects responded to strangers who had written a combination of attitudinal and neutral statements of fact (Byrne, Young, & Griffitt, in press), the relationship between proportion of positive reinforcements and attraction was found to be $Y = 5.36X + 5.06$. The present data, involving proportion of similar responses to personality items, are found not

TABLE 1

MEANS AND STANDARD DEVIATIONS OF ATTRACTION RESPONSES OF REPRESSERS, NEUTRALS, AND SENSITIZERS TOWARD STRANGERS DIFFERING IN PROPORTION OF SIMILAR RESPONSES TO R-S SCALE

R-S Scale level of Ss	Proportion of similar responses made by stranger							
	.20		.30		.80		Total	
	M	SD	M	SD	M	SD	M	SD
Repressers	5.06	2.48	7.50	2.29	8.67	2.30	6.96	2.81
Neutrals	6.24	2.16	8.37	2.48	10.33	2.16	8.35	2.81
Sensitizers	7.22	2.12	8.94	2.15	10.00	1.88	8.58	2.36
Total	6.17	2.43	8.29	2.39	9.70	2.25		

to depart significantly from that predicted by the above formula ($F = 1.84$, $df = 3/148$, ns).

The analysis of variance also indicated unexpected personality differences in that repressers gave the most negative attraction responses and sensitizers the most positive ones.

Discussion

At least with respect to the repression-sensitization dimension and within the limits of the present design, there is clear evidence that the relationship between personality similarity and attraction is a rectilinear one. In fact, proportion of similar personality responses influences attraction in precisely the same fashion as does proportion of similar attitude responses.

How may we account for the other significant finding? The most obvious possibility is that repression-sensitization influences attraction either in terms of pervasive differences in the characteristic level of need for affiliation or with respect to differences specifically in response to similarity and differences in

others. If the dissimilar responses of the stranger represent some degree of threat to the subject, the response might be analogous to repressor-sensitizer differences found elsewhere. A number of investigators report that threatening situations evoke more verbalized anxiety in sensitizers than in repressers (e.g., Davison, 1963; Lazarus & Alfert, 1964; Lomont, 1965; Pomeranz, 1963), but there is some evidence that such situations evoke more hostility in repressers than in sensitizers (e.g., Byrne & Sheffield, 1965) even though they attribute less hostility to themselves (e.g., Altrocchi, Shrauger, & McLeod, 1964). It is conceivable that it is the threat-hostility pattern in repressers which is responsible for their relatively more negative response to the strangers. If so, it would follow that these same repressor-sensitizer differences would be found in responses to attitudinal differences in strangers. A second experiment provides a test of that hypothesis.

EXPERIMENT II: RESPONSE OF REPRESSERS, NEUTRALS, AND SENSITIZERS TO SIMILARITY IN ATTITUDES

It was hypothesized that attraction responses toward strangers differing in attitude similarity-dissimilarity are a positive function of sensitizing as opposed to repressing defenses.

Method

The regular 182-item R-S Scale was administered to over 400 male and female students enrolled in a section of the introductory psychology course at the University of Texas. On the basis of test scores, three groups of subjects were selected: 49 sensitizers (scores of 70-115), 51 neutrals (scores of 40-48), and 49 repressers (scores of 6-29). The cutting points

TABLE 2

ANALYSIS OF VARIANCE OF ATTRACTION RESPONSES OF REPRESSERS, NEUTRALS, AND SENSITIZERS TOWARD STRANGERS DIFFERING IN PROPORTION OF SIMILAR RESPONSES TO R-S SCALE

Source	df	Adjusted MS	F
Response similarity	2	156.56	29.26***
Repression-sensitization	2	37.55	7.02**
Interaction	4	2.24 ^a	.42
Within	142	5.35	

** $p < .01$.

*** $p < .001$.

TABLE 3

MEANS AND STANDARD DEVIATIONS OF ATTRACTION RESPONSES OF REPRESSERS, NEUTRALS, AND SENSITIZERS TOWARD STRANGERS DIFFERING IN PROPORTION OF SIMILAR ATTITUDES

R-S Scale level of %	Proportion of similar attitudes held by stranger							
	.00		.50		1.00		1=100	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Repressers	6.47	2.12	8.00	1.28	11.62	1.87	8.65	2.81
Neutrals	6.61	2.45	9.13	2.19	11.89	2.31	9.22	3.21
Sensitizers	5.14	2.33	8.05	3.06	12.20	.65	8.49	3.60
Total	6.14	2.39	8.35	2.42	11.90	1.81		

for the personality groups differed somewhat from the first investigation primarily because the present group yielded higher scores on the R-S Scale, possibly as a function of differences attributable to the 127-item versus the 182-item forms.

At another class session, a 12-item attitude scale was administered. Each item consisted of a 6-point scale dealing with such issues as integration, political parties, necking and petting, and strict discipline for children. Subjects in each personality group were randomly assigned to one of three experimental conditions varying in proportion of similar attitudes attributed to a bogus stranger: .00, .50, and 1.00. In the .50 condition, the specific items of similarity and dissimilarity were selected by means of a table of random numbers and were different for each subject. As in previous attitude studies, similarity and dissimilarity are defined by responses falling on the same or on the opposite side of the neutral point with respect to the subject's response.

The experimental session followed essentially the same procedure as described in the previous investigation. In small groups, subjects were informed of the "interpersonal judgment" task, were given the attitude scale of an anonymous "stranger," and were asked to evaluate the stranger on the Interpersonal Judgment Scale.

Results

The means and standard deviations of the attraction responses are presented in Table 3. A 3×3 factorial analysis of variance, corrected for disproportionality (Wert et al., 1954), indicated a highly significant similarity-dissimilarity effect ($F = 82.18$, $df = 2/140$, $p < .001$). Neither the personality variable nor the interaction significantly affected attraction. The hypothesis was not confirmed.

Discussion

The explanation for the repressor-sensitizer differences in Experiment I in terms of differential response to similarity-dissimilarity does not appear to be an accurate one.

How else might the obtained differences be explained?

There is an additional source of variance in the stimuli presented to the subjects which has not yet been discussed. In preparing the R-S Scale of each stranger, different random patterns of items were prepared for each subject within a given level of similarity. With such a procedure, the repression-sensitization score of each stranger varies widely within each cell. It seemed possible that (a) the discrepancy between subject and stranger on the repression-sensitization dimension might fortuitously have been different for repressers, neutrals, and sensitizers, and (b) this discrepancy might influence attraction in addition to the effects of proportion of similar item responses.

The absolute R-S Scale discrepancy score² for each subject of Experiment I was determined, and it was found that the mean for repressers was 47.59, for neutrals 21.24, and for sensitizers 10.94. A simple one-way analysis of variance indicated that these differences are highly significant ($F = 56.88$, $df = 2/148$, $p < .001$). Thus, the first proposition was confirmed: the discrepancy differences paralleled the repressor-sensitizer differences. This

² Since a stranger with a given magnitude of discrepancy could be either more sensitizing or more repressing than the subject, the problem of the direction of the subject-stranger discrepancy was investigated before testing these propositions. For the 46 subjects (representing repressers, neutrals, and sensitizers) who could be approximately matched in terms of magnitude of plus or minus discrepancy with the stranger, a 2×3 analysis of variance was carried out with the 2 directions of discrepancy versus the 3 experimental conditions. The direction of discrepancy did not significantly affect attraction ($F = 3.76$, $df = 1/40$, ns).

TABLE 4

COMPARISON OF MEAN PREDICTED AND OBTAINED ATTRACTION RESPONSES IN EXPERIMENT I

R-S Scale level of Ss	Proportion of similar responses made by stranger					
	.20		.50		.80	
	Predicted	Obtained	Predicted	Obtained	Predicted	Obtained
Repressors	4.96	5.06	7.22	7.50	9.60	8.67
Neutrals	6.61	6.24	8.16	8.37	9.96	10.33
Sensitizers	7.23	7.22	8.74	8.94	10.08	10.00

suggests the possibility that the findings with respect to personality differences were artifactual.

To test the effect of discrepancy on attraction, a correlational analysis was employed. For the total group of 151 students, the subject-stranger discrepancy score correlated $-.52$ ($p < .001$) with attraction. To control for the confounding effects of experimental conditions, correlations were also obtained separately within each condition, yielding coefficients of $-.39$ ($p < .01$) in the .20 condition, $-.28$ ($p < .05$) in the .50 condition, and $-.34$ ($p < .05$) in the .80 condition. It seems that subjects are responding to R-S Scale discrepancy as well as to proportion of similar item responses. The two stimulus variables are not completely independent because discrepancy tends to decrease as proportion of similar items increases ($r = -.56$, $p < .001$). Nevertheless, each stimulus dimension does influence attraction even with the other dimension controlled. The correlation between proportion of similar items and attraction with R-S Scale discrepancy partialled out is $.33$ ($p < .001$), and the correlation between R-S Scale discrepancy and attraction with proportion of similar items partialled out is $-.33$ ($p < .001$). In order to provide more conclusive evidence that it is subject-stranger discrepancy rather than subject repression-sensitization which accounts for these findings, it was necessary to utilize double partial correlations. The relationship between R-S Scale discrepancy and attraction with *both* proportion of similar items and subject's R-S Scale score held constant is a significant one ($r_{12.34} = .20$, $p < .02$), while the analogous relationship between subject's R-S Scale score

and attraction is not significantly different from zero ($r_{12.34} = .11$, *ns*).

The situation, then, is one in which the dependent variable is affected by two independent variables, one of which was varied as part of the experimental design and one of which was accidentally manipulated. It is possible, therefore, to conceptualize attraction as a function of both variables, each of which represents a different aspect of personality similarity. One way of describing this relationship is by means of a multiple-correlation coefficient which is found to equal $.59$ ($p < .001$).

Returning to the original problem of accounting for repressor-sensitizer differences in attraction, it would seem that they may be explained simply in terms of group differences in discrepancy scores. This can be demonstrated rather convincingly by predicting each subject's attraction response on the basis of a multiple-regression equation ($X_1' = -.04X_2 + 3.88X_3 + 7.16$). In Table 4, the attraction scores as predicted by this equation are presented along with the obtained attraction scores. The obviously close correspondence is confirmed by the fact that the mean difference between predicted and obtained responses was found to be $-.004$, yielding a t of $.02$ ($df = 150$, *ns*).

The methodological approach utilized here confirmed the proposed positive relationship between personality similarity and attraction and has shown that subjects are surprisingly sensitive to the similarity cues presented by the stranger. They respond not only to specific response similarity, but also to similarity at a more abstract or generalized level, that is, to the personality dimension itself. While

the problem of awareness has not been explored in the attraction studies, it appears unlikely that the typical subject would be able to verbalize the degree to which he and a stranger responded in the same or opposite ways on 127 T-F items. It appears even less likely that the typical subject could verbalize the discrepancy dimension when the dimension itself, the scoring system, and both his and the stranger's scores are all unknown to him. Nevertheless, attraction is responsive to variation in both stimulus dimensions, and the nature of the relationship is entirely consistent with the findings in the attitude studies.

REFERENCES

- ALTROCCHI, J. Dominance as a factor in interpersonal choice and perception. *Journal of Abnormal and Social Psychology*, 1959, 59, 303-308.
- ALTROCCHI, J., SHRAUGER, S., & MCLEOD, M. A. Attribution of hostility to self and others by expressors, sensitizers, and repressors. *Journal of Clinical Psychology*, 1964, 20, 233.
- BANTA, T. J., & HETHERINGTON, M. Relations between needs of friends and fiancés. *Journal of Abnormal and Social Psychology*, 1963, 66, 401-404.
- BECKER, G. The complementary-needs hypothesis, authoritarianism, dominance, and other Edwards Personal Preference Schedule scores. *Journal of Personality*, 1964, 32, 45-56.
- BONNEY, M. E. A sociometric study of the relationship of some factors to mutual friendships on the elementary, secondary, and college levels. *Sociometry*, 1946, 9, 21-47.
- BOWERMAN, C. E., & DAY, B. R. A test of the theory of complementary needs as applied to couples during courtship. *American Sociological Review*, 1956, 21, 602-605.
- BRIM, O. G., JR., & HOFF, D. B. Individual and situational differences in desire for certainty. *Journal of Abnormal and Social Psychology*, 1957, 54, 225-229.
- BROCK, T. C. Communicator-recipient similarity and decision change. *Journal of Personality and Social Psychology*, 1965, 1, 650-654.
- BYRNE, D. Interpersonal attraction and attitude similarity. *Journal of Abnormal and Social Psychology*, 1961, 62, 713-715.
- BYRNE, D. Repression-sensitization as a dimension of personality. In B. A. Maher (Ed.), *Progress in experimental personality research*. Vol. 1. New York: Academic Press, 1964. Pp. 169-220.
- BYRNE, D., BARRY, J., & NELSON, D. Relation of the revised Repression-Sensitization Scale to measures of self-description. *Psychological Reports*, 1963, 13, 323-334.
- BYRNE, D., & BUEHLER, J. A. A note on the influence of propinquity upon acquaintanceships. *Journal of Abnormal and Social Psychology*, 1955, 51, 147-148.
- BYRNE, D., & CLORE, G. L., JR. Predicting interpersonal attraction toward strangers presented in three different stimulus modes. *Psychonomic Science*, 1966, 4, 239-240.
- BYRNE, D., CLORE, G. L., JR., & WORCHEL, P. Effect of economic similarity-dissimilarity on interpersonal attraction. *Journal of Personality and Social Psychology*, 1966, 4, 220-224.
- BYRNE, D., & GRIFFITT, W. A developmental investigation of the law of attraction. *Journal of Personality and Social Psychology*, 1966, 4, 699-703.
- BYRNE, D., & NELSON, D. Attraction as a linear function of proportion of positive reinforcements. *Journal of Personality and Social Psychology*, 1965, 1, 659-663.
- BYRNE, D., NELSON, D., & REEVES, K. Effects of consensual validation and invalidation on attraction as a function of verifiability. *Journal of Experimental Social Psychology*, 1966, 2, 98-107.
- BYRNE, D., & SHEFFIELD, J. Response to sexually arousing stimuli as a function of repressing and sensitizing defenses. *Journal of Abnormal Psychology*, 1965, 70, 114-118.
- BYRNE, D., YOUNG, R. K., & GRIFFITT, W. The reinforcement properties of attitude statements. *Journal of Experimental Research in Personality*, in press.
- COHEN, A. R. Experimental effects of ego-defense preference on interpersonal relations. *Journal of Abnormal and Social Psychology*, 1956, 52, 19-27.
- COHEN, A., STOTLAND, E., & WOLFE, D. M. An experimental investigation of need for cognition. *Journal of Abnormal and Social Psychology*, 1955, 51, 291-297.
- CORSINI, R. J. Understanding and similarity in marriage. *Journal of Abnormal and Social Psychology*, 1956, 52, 327-332.
- DAVISON, L. A. Adaptation to a threatening stimulus. Unpublished doctoral dissertation, University of California, Berkeley, 1963.
- DEUTSCH, M., & SOLOMON, L. Reactions to evaluations by others as influenced by self-evaluations. *Sociometry*, 1959, 22, 93-112.
- FESTINGER, L. Informal social communication. *Psychological Review*, 1950, 57, 271-282.
- FESTINGER, L. A theory of social comparison processes. *Human Relations*, 1954, 7, 117-140.
- FESTINGER, L., SCHACHTER, S., & BACK, K. *Social pressures in informal groups: A study of human factors in housing*. New York: Harper, 1950.
- GOLIGHTLY, C., & BYRNE, D. Attitude statements as positive and negative reinforcements. *Science*, 1964, 146, 798-799.
- GORDON, J. E. Interpersonal predictions of repressors and sensitizers. *Journal of Personality*, 1957, 25, 686-698.
- GRIFFITT, W. B. Interpersonal attraction as a function of self-concept and personality similarity-dissimilarity. *Journal of Personality and Social Psychology*, 1966, 4, 581-584.

- HOFFMAN, L. R. Similarity of personality: A basis for interpersonal attraction? *Sociometry*, 1958, 21, 300-308.
- HOFFMAN, L. R., & MAIER, N. R. F. An experimental reexamination of the similarity-attraction hypothesis. *Journal of Personality and Social Psychology*, 1966, 3, 145-152.
- IZARD, C. E. Personality similarity and friendship. *Journal of Abnormal and Social Psychology*, 1960, 61, 47-51. (a)
- IZARD, C. E. Personality similarity, positive affect, and interpersonal attraction. *Journal of Abnormal and Social Psychology*, 1960, 61, 484-485. (b)
- IZARD, C. E. Personality similarity and friendship: A follow-up study. *Journal of Abnormal and Social Psychology*, 1963, 66, 598-600.
- JONES, E. E. *Ingratiation: A social psychological analysis*. New York: Appleton-Century-Crofts, 1964.
- KATZ, I., COHEN, M., & CASTIGLIONE, L. Effect of one type of need complementarity on marriage partners' conformity to one another's judgments. *Journal of Abnormal and Social Psychology*, 1963, 67, 8-14.
- KATZ, I., GLUCKSBERG, S., & KRAUSS, R. Need satisfaction and Edwards PPS scores in married couples. *Journal of Consulting Psychology*, 1960, 24, 205-208.
- KERCKHOFF, A. C., & DAVIS, K. E. Value consensus and need complementarity in mate selection. *American Sociological Review*, 1962, 27, 295-303.
- LAZARUS, R. S., & ALFERT, E. Short-circuiting of threat by experimentally altering cognitive appraisal. *Journal of Abnormal and Social Psychology*, 1964, 69, 195-205.
- LEVINGER, G. Note on need complementarity in marriage. *Psychological Bulletin*, 1964, 61, 153-157.
- LEVINGER, G., & BREEDLOVE, J. Interpersonal attraction and agreement: A study of marriage partners. *Journal of Personality and Social Psychology*, 1966, 3, 367-372.
- LOMONT, J. F. The repression-sensitization dimension in relation to anxiety responses. *Journal of Consulting Psychology*, 1965, 29, 84-86.
- LOTT, B. E., & LOTT, A. J. The formation of positive attitudes toward group members. *Journal of Abnormal and Social Psychology*, 1960, 61, 297-300.
- LUNDY, R. M. Self-perceptions regarding masculinity-femininity and descriptions of same and opposite sex sociometric choices. *Sociometry*, 1958, 21, 238-246.
- MAISONNEUVE, J. A contribution to the sociometry of mutual choices. *Sociometry*, 1954, 17, 33-46.
- MCDONALD, R. D. The effect of reward-punishment and affiliation need on interpersonal attraction. Unpublished doctoral dissertation, University of Texas, 1962.
- MEHLMAN, B. Similarity in friendships. *Journal of Social Psychology*, 1962, 57, 195-202.
- MILLER, N., CAMPBELL, D. T., TWEDT, H., & O'CONNELL, E. J. Similarity, contrast, and complementarity in friendship choice. *Journal of Personality and Social Psychology*, 1966, 3, 3-12.
- MORTON, A. Similarity as a determinant of friendship: A multidimensional study. *Dissertation Abstracts*, 1960, 20, 3857-3858.
- MURSTEIN, B. I. The complementary need hypothesis in newlyweds and middle-aged married couples. *Journal of Abnormal and Social Psychology*, 1961, 63, 194-197.
- NELSON, D. The effect of differential magnitude of reinforcement on interpersonal attraction. Unpublished doctoral dissertation, University of Texas, 1965.
- NEWCOMB, T. M. The prediction of interpersonal attraction. *American Psychologist*, 1956, 11, 575-586.
- PEPITONE, A. *Attraction and hostility*. New York: Atherton, 1964.
- PERVIN, L. A. The need to predict and control under conditions of threat. *Journal of Personality*, 1963, 31, 570-587.
- PINTNER, R., FORLANO, G., & FREEDMAN, H. Personality and attitudinal similarity among classroom friends. *Journal of Applied Psychology*, 1937, 21, 48-65.
- POMERANZ, D. M. The repression-sensitization dimension and reactions to stress. Unpublished doctoral dissertation, University of Rochester, 1963.
- REILLY, M. S. A., COMMINIS, W. D., & STEFIC, E. C. The complementarity of personality needs in friendship choice. *Journal of Abnormal and Social Psychology*, 1960, 61, 292-294.
- ROSENFELD, H. M., & JACKSON, J. Temporal mediation of the similarity-attraction hypothesis. *Journal of Personality*, 1965, 33, 649-656.
- RYCHLAK, J. F. The similarity, compatibility, or incompatibility of needs in interpersonal selection. *Journal of Personality and Social Psychology*, 1965, 2, 334-340.
- SECORD, P. F., & BACKMAN, C. W. Interpersonal congruency, perceived similarity, and friendship. *Sociometry*, 1964, 27, 115-127.
- THORPE, J. G. A study of some factors in friendship formation. *Sociometry*, 1955, 18, 207-214.
- VAN DYNE, V. E. Personality traits and friendship formation in adolescent girls. *Journal of Social Psychology*, 1940, 12, 291-303.
- WERT, J. E., NEDT, C. O., & AHMANN, J. S. *Statistical methods in educational and psychological research*. New York: Appleton-Century-Crofts, 1954.
- WINCE, R. F., KTSANES, T., & KTSANES, V. Empirical elaboration of the theory of complementary needs in mate-selection. *Journal of Abnormal and Social Psychology*, 1955, 51, 508-513.
- ZANDER, A., & HAVELIN, A. Social comparison and interpersonal attraction. *Human Relations*, 1960, 13, 21-32.
- ZIMBARDO, P., & FORMICA, R. Emotional comparison and self-esteem as determinants of affiliation. *Journal of Personality*, 1963, 31, 141-162.

(Received April 1, 1966)

EFFECTS OF FAMILIARIZATION AND GROUP DISCUSSION UPON RISK TAKING¹

JAMES P. FLANDERS

AND

DONALD L. THISTLETHWAITE

Vanderbilt University

Vanderbilt University

Results of a 2×2 factorial experiment with 180 male college students replicated previous findings concerning the main effects of the independent variables of group discussion and familiarization upon shifts in risk-taking dispositions and revealed a significant interaction between these variables, indicating that group discussion produces risk-taking shifts among unfamiliarized Ss but has no effect upon familiarized Ss. Predictions concerning this interaction based upon previously elaborated explanations of the risky shift effect in terms of group processes (responsibility diffusion and interpersonal influence) were unconfirmed. The observed risky shifts can be interpreted in terms of a process of increased comprehension, which is theorized to be the outcome of interpolated familiarization or discussion procedures.

Several experiments using pretest-posttest control-group designs have demonstrated a "risky shift" effect, in which subjects advocated riskier solutions to risk-taking problems after discussing the problems than they did before the discussion. In the initial experiments (Bem, Wallach, & Kogan, 1965; Stoner, 1961; Wallach, Kogan, & Bem, 1962, 1964) the discussion treatments which produced sizable risky shifts were invariably accompanied by unanimous group decisions. Recently, Wallach and Kogan (1965) found that groups exposed to discussion without the requirement of consensus exhibited risky shifts fully as great as those in a condition requiring discussion to consensus. They interpreted these results as indicating that group discussion is both a necessary and sufficient condition for generating the risky shift effect. Wallach and Kogan proposed a responsibility diffusion interpretation, which states that direct verbal confrontation provided by group discussion leads to the development of affective interdependencies, which in turn mediates a process of responsibility diffusion. According to this interpretation the increased willingness to take risks which follows group

discussion is produced by a decreased feeling of personal responsibility.

A rival explanation, which may be called a comprehension interpretation, is suggested by Bateson's (1966) recent experiment. In this experiment similar risky shifts were obtained in a condition providing interpolated individual study of the risk-taking problems instead of group discussion. Moreover, the risky shift effect obtained in such a familiarization condition was fully as great as that obtained in a condition requiring discussion to consensus. According to Bateson, the effective ingredient in both discussion and familiarization treatments may be the more thorough consideration and comprehension of the risk-taking problems which additional study provides. Presumably better comprehension of the relevant information in each test item, or of the alternative probabilities from which choices are to be made, leads to reduced cautiousness (increased willingness to advocate risky solutions).

Since Bateson's experiment did not include a condition in which subjects were exposed to both familiarization and group discussion, it is impossible to say whether the responsibility diffusion and comprehension explanations should be considered mutually exclusive. One of the main purposes of this experiment was to explore the compatibility of the responsibility diffusion and comprehension interpretations by examining the interaction of the independent variables of familiarization

¹ This research was partially supported by a Public Health Service Predoctoral Fellowship (NH-32,672) awarded to the senior author and is based upon a master's thesis submitted to Vanderbilt University. The authors are indebted to Norman E. Wheeler for facilitating the computer analysis of the results.

and group discussion. Depending upon the theoretical processes assumed, varying predictions concerning this interaction may be made.

Let us suppose first that (a) group discussion produces responsibility diffusion, (b) familiarization produces improved comprehension, and (c) both theoretical processes (responsibility diffusion and improved comprehension) contribute to the production of the risky shift effect. To predict the presence or absence of an interaction we must of course make more specific assumptions about the joint operation of these processes. If we assume that the operation of one of these processes does not affect the operation of the other, then it follows that the joint effect of the two independent variables should reflect a simple additivity of the main effects. On the other hand, if it is assumed that the two processes are complementary and mutually facilitative, then it follows that the joint effect of the two independent variables should be an enhanced risky shift significantly greater than would be produced from a simple addition of the main effects.

An alternative set of assumptions, postulating only a single theoretical process, might include the following: (a) Group discussion produces improved comprehension; (b) familiarization produces improved comprehension; and (c) improved comprehension is the sole process contributing to the risky shift effect. As before, in order to make specific predictions about the interaction effect, we must be more specific about the hypothesized process when both independent variables are operative. If we assume that the level of comprehension reaches an asymptote with only moderate levels of familiarization or of group discussion, then it follows that the joint effect of both independent variables should be no greater than the effect of one variable acting alone. But if we assume that the level of comprehension reaches an asymptote only after extensive levels of familiarization or of group discussion, then it follows that the joint effects of both variables should consist of a simple additivity of main effects.

A second purpose of the present investigation was to provide data for evaluating a third hypothetical process which may be in-

volved in the risky shift effect: namely, high risk takers may more readily take the initiative in social situations and exert more influence in group discussions, so that the greater interpersonal influence of high risk takers is the cause of the group's movement toward greater risk taking. First, it may be noted that if we substitute such a process of differential interpersonal influence for responsibility diffusion in the first set of assumptions outlined above, then the initial expectations concerning interaction effects follow. Second, this study sought to provide a test of the related hypothesis that perceived influence in group discussions is positively correlated with initial risk-taking dispositions. In fact, Wallach et al. (1962) found these two variables to be significantly correlated even when judgments of popularity were held constant. The present experiment also tested the related hypotheses that extraversion would be positively correlated with both initial risk level and perceived influence.

It was hoped that with knowledge of the interaction effect between the independent variables and of the related correlations between risk taking and other variables it would be possible to determine which set of assumptions offers the best overall predictions of risky shift effects under a variety of conditions.

METHOD

Pretest Procedures

Three subjects were scheduled to appear in the experimental room at a given time. Subjects were seated about a round table, on which partitions were placed to separate each subject from the view of the others, and given the same instructions employed by Wallach and Kogan (1965) for completing the risk-taking dilemmas. These risk-taking measures consisted of a set of 12 real-life dilemmas first devised by Wallach and Kogan (1959, 1961) and reproduced in full by Kogan and Wallach (1964, Appendix E). Each of the dilemmas describes a situation in which a protagonist is faced with a choice between two actions, X and Y, whose outcomes differ in attractiveness and probability of occurrence. The successful outcome associated with Alternative X is always more attractive than the successful outcome associated with Alternative Y, but the probability of attaining the former is less than that of attaining the latter. Subjects are asked to consider themselves as advisors to the protagonist in each story, and to indicate the minimum

probability of success they would consider acceptable in order to recommend the Action X which potentially leads to the more attractive outcome. The probabilities listed for each of the 12 dilemmas include 1, 3, 5, 7, and 9 chances out of 10; in addition each item includes a category (scored 10) in which the respondent may indicate his belief that the protagonist should not take Action X "no matter what the probabilities." Risk-taking scores were computed for a given measurement occasion by adding the chosen probabilities for each item and multiplying by 10. Thus, a subject who chose 5/10 on each of the 12 dilemmas would have a risk-taking score of $(12) \times (5/10) \times (10) = 60$. Low scores reflected a greater willingness to advocate risky solutions.

Eysenck's (1956) 24-item extraversion scale was administered to all subjects as the last part of the pretest. Following administration of the personality measure, the procedures for the different experimental conditions varied as indicated below.

Experimental Manipulations

A 2×2 factorial design was used in which the independent variables of familiarization and discussion with consensus were manipulated. Instructions used in each of the four cells of the design follow.

Familiarization without discussion. After collecting the pretest materials, the experimenter handed out new blank copies of the dilemma booklets and said:

You will now *prepare* for a group discussion which will involve all three of you. You will discuss each of the 12 situations. Your task now is to restudy each choice, with the view of making certain that you have not overlooked some relevant information. Try to clarify for yourself the nature of each choice open to the central person and the character of the odds attached to each alternative. We want each of you to be prepared to discuss the choices to be made so that you will not have to spend group discussion time restudying the problem. Also, the discussion should then concentrate only on the important issues relevant to each choice. It is strongly recommended that in the 25 minutes you now have to restudy that you note in the questionnaire booklet what you consider to be the important issues. Use whatever method you use every day, such as underlining, starring, jotting in the margin, or making lists of pros and cons. You will be allowed to refer to your markings in this booklet during the discussion. Feel free to read between the lines. You have 25 minutes.

During the period of individual study of the dilemmas, the partitions remained in place on the table, and none of the subjects communicated with each other. The experimenter next gave the following instructions for the posttest measure of risk taking:

Before proceeding to the group discussion, we will have a short bit of further individual work. I want you to go back over each of these situations and indicate your own present personal

decision with a "P." It is quite natural that some further thoughts have occurred to you since you indicated your reactions to the situations the first time. You need not consider yourself bound by any of the past decisions, because your task now is simply to indicate the recommendation you would give to the central figure *at the present time*. Remember, you are to indicate your own personal decision at the present time with a P.

Since subjects in this condition anticipated a discussion but did not actually participate in one, the experimenter provided at the conclusion of the session a debriefing which explained the need for deception and requested subjects not to discuss the experiment with others.

Discussion without familiarization. Having collected the pretest materials, the experimenter removed the partitions separating subjects and distributed new blank copies of the booklets. He proceeded by saying:

We are now ready to begin the group discussion. What we are really interested in *now* is having the group discuss every situation in turn and arrive at a unanimous decision on each. You will recognize that your unanimous decision is different from a majority vote, by the way. You all must agree. This time you may not return to a question; discuss each one until the group decision is reached and then go on to the next. When the group reaches its decision, you are to mark it on the questionnaire you have so that you have a record of the group's decisions. I am not going to participate in the discussion, although I will be here to answer any procedural questions which may arise. Remember that it is only the group's final decision in which we are interested.

After each group had discussed—and reached a unanimous consensus with respect to—each of the dilemmas, the experimenter gave the instructions for the posttest measure of risk taking. These instructions were virtually identical for all conditions. The only variation was that in the two experimental conditions involving group discussion, the experimenter omitted the phrase "before proceeding to the group discussion" from the second paragraph of instructions for the immediately preceding condition.

Familiarization and discussion. Specific instructions for this condition duplicated the first paragraph of instructions for the familiarization without discussion condition, and the complete set of instructions for the immediately preceding discussion without familiarization condition.

Neither familiarization nor discussion. Following administration of the personality measure, the experimenter proceeded immediately to the instructions for the posttest measure of risk taking described above for the familiarization without discussion condition.

Influence Measures

Participants in the group discussions described above were asked, as part of an exercise following

the posttest, to rank order the group members (including themselves) with respect to overall "degree of influence" in the discussion. A subject's perceived influence score was the sum of the three ranks assigned to him; a low sum indicated greater perceived influence.

Subjects

Subjects were 180 male undergraduates at Vanderbilt University enrolled in the introductory psychology course. All subjects volunteered for this experiment, and each group of three subjects volunteering for a given experimental session was randomly assigned to one of the four experimental conditions. Fifteen groups of three persons each were assigned to each condition.

RESULTS

In the analysis of results, a group's risk-taking score, whether based on the pretest or posttest measure, was the arithmetic sum of the risk-taking score of its three members. Shift scores for each group were obtained by subtracting pretest from posttest scores.

An F test comparing the mean pretest scores of groups in the four experimental conditions was insignificant ($F < 1$); thus the random assignment of groups to the conditions was effective in ruling out initial differences in risk-taking dispositions.

An analysis of variance was performed on the group shift scores and is presented in Table 1. It reveals that there was a highly significant main effect attributable to the familiarization manipulation. Also there was a significant interaction between the familiarization and group-discussion variables. The group-discussion manipulation did not have a significant main effect.

Table 2 contains average group risk-taking and shift scores for each experimental condition and presents t tests for evaluating

TABLE 1
ANALYSIS OF VARIANCE OF SHIFT SCORES OF
GROUPS IN EXPERIMENTAL CONDITIONS

Source	df	MS	F
Familiarization (A)	1	1025.07	8.56****
Discussion (B)	1	232.07	1.94
A \times B	1	777.60	6.49**
Error	56	119.74	
Total	59	148.14	

** $p < .02$.

**** $p < .005$.

whether each of the mean shifts was significantly different from zero. The mean shift in riskiness was significantly greater than zero in Conditions F (familiarization without discussion), FD (familiarization and discussion), and D (discussion without familiarization). The slight increase in conservatism shown by Condition N (neither familiarization nor discussion) was not significantly different from zero.

Risk-taking shifts among the four experimental conditions are compared in Table 3. Conditions, F, FD, and D each exhibited greater risky shifts than Condition N, but they did not differ significantly from each other.

Correlations between initial risk-taking scores and other variables were based on the scores of the 90 subjects participating in group discussions. As expected, initial risk-taking scores had a significant correlation with perceived influence scores ($r = .30$), but—contrary to the hypothesis—had only an insignificant correlation with extraversion scores. Also the correlation between extraversion and perceived influence scores was insignificant.

Finally, an analysis of the items making up

TABLE 2
SHIFTS IN RISK TAKING UNDER EACH CONDITION

Condition	Mean pretest score	Mean posttest score	Mean shift*	t
F-Familiarization without discussion	216.3	202.4	-13.9	5.38***
FD-Familiarization and discussion	220.3	209.6	-10.7	3.36***
D-Discussion without familiarization	209.7	200.1	-9.6	3.40***
N-Neither familiarization nor discussion	218.1	219.6	+ 1.5	.57

Note.—N = 15 groups within each condition.

* Negative sign indicates increase in riskiness.

*** $p < .001$ (2-tailed).

TABLE 3

COMPARISON OF RISK-TAKING SHIFTS BETWEEN
PAIRS OF CONDITIONS

Comparison	<i>t</i>
F vs. N	4.15***
FD vs. N	2.94**
D _a vs. N	2.86**
F vs. D	1.13
F vs. FD	.80
FD vs. D	.25

** $p < .02$ (2-tailed).*** $p < .001$ (2-tailed).

the dilemma test was performed to investigate the internal homogeneity of this risk-taking measure. Using the risk-taking data, for all 180 subjects an alpha coefficient of .61 was found for the pretest scores, while a coefficient of .59 was obtained for the posttest scores. Correlations between each item response and the individual risk-taking scores based upon all 12 dilemmas were computed from the pretest data; these correlations ranged from .52 to .33. Using the numbering of the dilemmas described by Kogan and Wallach (1964, Appendix E), the eighth and twelfth had the lowest item total-score correlations and were the only two on which the 180 subjects exhibited an average shift score which was positive (i.e., in the conservative direction).

DISCUSSION

The present experimental results indicate that discussion to consensus has the effect of producing a risky shift among subjects who have *not* had interpolated familiarization with the risk-taking problems. However, among subjects who have been given the opportunity of further individual study of the risk-taking problems, discussion to consensus has no effect upon risk-taking scores.

The responsibility-diffusion interpretation can account for the former, but not for the latter finding. These findings concerning the interaction of the two variables of familiarization and group discussion provide a more persuasive confirmation of the responsibility-diffusion hypothesis than Bateson's (1966) finding—which was confirmed by the present comparisons—that a familiarization without discussion condition was as effective in pro-

ducing a risky shift as a discussion without familiarization condition. One might maintain, in attempting to account for the latter result, that responsibility diffusion may be one of many sufficient causes of the risky shift and that improved comprehension of the risk-taking tasks may be another. However, the present findings pose a particularly difficult explanatory problem for the responsibility-diffusion explanation. It cannot explain why group discussion to a consensus, which should increase affective interdependencies and responsibility diffusion in "familiarized" as well as "unfamiliarized" subjects, failed to have any effect upon the risk-taking scores of subjects who first had the opportunity of additional study of the risk-taking problems. The possibility that the particular discussion procedures used were ineffectual seems to be ruled out since discussion did produce a significant risky shift effect among subjects not exposed to a familiarization procedure.

Bateson (1966) acknowledged, in attempting to reconcile his findings with those of Wallach and Kogan (1965), that since he had employed only 5 of the 12 dilemmas his groups remained in existence a shorter period of time, and therefore may not have developed the same degree of affective interdependency as groups which consider the entire set of dilemmas. The present experiment seems to rule out this possibility, however, since it employed the entire set of 12 dilemmas and included instructions which were virtually identical to those used in the discussion and consensus condition of Wallach and Kogan (1965). Moreover, the present item analysis of the 12 dilemmas revealed a pattern of risky shifts on individual items quite similar to that found previously: Wallach et al. (1962) found that the average shift scores for both men and women were negative (in a risky direction) for 10 of the 12 dilemmas; furthermore, both studies found that the average shift score was positive (in a conservative direction) for the twelfth item in the set. The chief difference was that the latter investigators found an overall conservative shift on the fifth item rather than the eighth item. It is of interest to note that the fifth, eighth, and twelfth items were not included among the five employed by Bateson

(1966), so that his results cannot be attributed to an unrepresentative selection of test items.

In order to determine whether there was any evidence in the present experiment suggesting that influence attempts operate to bring opinion deviates into closer conformity with a group norm, an additional analysis was made of the changes in within-group variability of risk-taking scores among group members under each of the experimental conditions. Wallach and Kogan (1965) found that the two discussion conditions which produced the greatest shifts in risk taking also produced marked reductions in within-group variability. For this analysis, we used a measure of within-group variability similar to that employed by Wallach and Kogan: within each group the range of the three individual post-test risk-taking scores was subtracted from the range of the three individual pretest risk-taking scores, yielding a single difference score (d) for each group of three subjects. A positive difference between these two ranges, of course, indicates a reduction in within-group variability. An analysis of variance of the d scores reveals a significant main effect associated with the discussion manipulation ($F = 31.20, p < .001$), but there is neither a significant main effect for the familiarization manipulation ($F < 1$) nor a significant interaction between the two independent variables ($F < 1$). Table 4 contains t tests for evaluation whether the shift in within-group variability was significantly different from zero in each condition.

These results essentially duplicate the findings of Wallach and Kogan (1965) and indicate that there was a significant reduction in within-group variability for Conditions D and

FD—the two conditions providing group discussion—but no significant reduction in variability for Conditions F and N—the two conditions in which group discussion was not present. In fact, Condition N showed a marginally significant *increase* in within-group variability. Separate comparisons indicate that Conditions D and FD did not differ significantly from each other in the degree of narrowing of intragroup heterogeneity, but that each differed significantly from the average shifts in variability in Conditions F and N. It seems clear that group-influence processes were at work in the groups which discussed the dilemmas.

However, these influence processes are generally unrelated to the risky shifts shown by the groups. For example, Condition F which produced the largest risky shift, showed an average increase, rather than a decrease, in within-group variability. As a further check on the possible role of group-influence processes in the production of the risky shift, a correlation was computed between the group risk-taking shift scores and the shifts in within-group variability (d scores) of the 60 groups; this correlation was low and insignificant ($r = -.11$). These comparisons indicate that the observed reductions in within-group variability played little or no role in the risky shift effects in the present experiment. Presumably, in the discussion groups influence processes are directed with more or less equal intensity toward deviates on both the risky and the conservative sides of the average initial risk-taking levels.

The foregoing results provide no support for the view that group processes mediate the risky shift effect. Consequently, the significant correlation between initial risk taking and perceived influence scores probably should not be interpreted as indicating that high risk takers exercise greater interpersonal influence in the group discussions. This correlation may be explained equally well by supposing that influence ranks were assigned on the basis of noting that the group's consensus usually most closely matched the initial position of the high risk taker in the group. In other words, the group's movement toward riskier solutions may have resulted in group members ascribing more influence to the initially high risk taker.

TABLE 4

SHIFTS IN WITHIN-GROUP RANGE OF RISK-TAKING LEVELS UNDER EACH CONDITION

Condition	Average shift (d)	SE_d	t
D	11.00	3.20	3.44***
FD	6.87	2.02	3.40***
N	- 4.33	1.84	2.35*
F	- 4.33	2.21	1.96

Note.—N = 15 groups within each condition.

* $p < .05$ (2-tailed).

*** $p < .001$ (2-tailed).

On the other hand, the interpretation suggested by Bateson (1966) can readily explain both the main and interaction effects observed in this experiment. Improved comprehension may be the outcome of a discussion without familiarization condition, just as it may be the outcome of a familiarization without discussion condition. When subjects are given the opportunity to more adequately grasp the relevant facts or choices posed by a hypothetical dilemma, they become less cautious and more willing to advocate risky solutions to the dilemmas. According to this interpretation, discussion produces risky shifts among unfamiliarized subjects because the discussion provides time for further study and understanding of the tasks. Moreover, discussion has no effect upon the risk-taking dispositions of familiarized subjects (who have already achieved an adequate comprehension of the task) simply because further study has little to add.

REFERENCES

- BATESON, N. Familiarization, "group discussion, and risk-taking. *Journal of Experimental Social Psychology*, 1966, 2, 119-129.
- BEM, D. J., WALLACH, M. A., & KOGAN, N. Group decision making under risk of aversive consequences. *Journal of Personality and Social Psychology*, 1965, 1, 453-460.
- EYSENCK, H. J. The inheritance of extraversion-introversion. *Acta Psychologica*, 1956, 12, 95-110.
- KOGAN, N., & WALLACH, M. A., *Risk taking: A study in cognition and personality*. New York: Holt, Rinehart & Winston, 1964.
- STONER, J. A. F. A comparison of individual and group decisions involving risk. Unpublished master's thesis, Massachusetts Institute of Technology, School of Industrial Management, 1961.
- WALLACH, M. A., & KOGAN, N. Sex differences and judgmental processes. *Journal of Personality*, 1959, 27, 555-564.
- WALLACH, M. A., & KOGAN, N. Aspects of judgment and decision making: Interrelation shifts and changes with age. *Behavioral Science*, 1961, 6, 23-36.
- WALLACH, M. A., & KOGAN, N. The roles of information, discussion, and consensus in group risk taking. *Journal of Experimental Social Psychology*, 1965, 1, 1-19.
- WALLACH, M. A., KOGAN, N., & BEM, D. J. Group influence on individual risk taking. *Journal of Abnormal and Social Psychology*, 1962, 65, 75-86.
- WALLACH, M. A., KOGAN, N., & BEM, D. J. Diffusion of responsibility and level of risk taking in groups. *Journal of Abnormal and Social Psychology*, 1964, 68, 263-274.

(Early publication received August 8, 1966)

BRIEF ARTICLES

PUPILLARY RESPONSE, CONDITIONING, AND PERSONALITY

DAVID S. HOLMES¹

Northwestern University

This research tested the relationship of levels of acetylcholine and norepinephrine (as inferred from speed of pupillary constriction and dilation) to awareness of an environmental contingency, performance on a verbal conditioning task, and personality. High levels of acetylcholine, inferred from rapid constriction, were found to be significantly related to greater awareness of an environmental contingency, superior performance in verbal conditioning, and to introverted personality traits as measured by self-report and peer ratings. Relationships between levels of norepinephrine, inferred from speed of dilation, awareness, conditioning, and personality were for the most part inconsistent and nonsignificant. The conclusion was that speed or efficiency of neural transmission at cholinergic synapses, where acetylcholine is the transmitter substance, was related to conditioning and personality. The lack of significant findings regarding norepinephrine was attributed to the probable limitation of this mediator to synapses of the autonomic system.

In the past few years a number of important findings have been reported relating brain chemistry to learning. Rosenzweig, Krech, and Bennett (1958, 1961) have shown that rats which performed better in three different types of situations had greater amounts of cholinesterase (ChE) in the cerebral cortex. It has been known for some time that acetylcholine (ACh) is important in facilitating neural transmission at the synapses (Eccles, 1957), and since ChE breaks up or hydrolyzes ACh after the transmission of the impulse, ChE was used as an indicator of the amount of ACh available; that is, better learning was accompanied by higher levels of ChE and, inferentially, higher levels of ACh. In another approach to the study of neural transmission and learning, Russell, Watson, and Frankenhaeuser (1961) have shown that animals fed a ChE inhibitor, Systox, which consequently reduced the hydrolysis of ACh, showed slower extinction than normal animals. All of these data suggest that there is an important relationship between speed or efficiency of neural transmission and learning.

The application of this work to humans was retarded by the lack of appropriate techniques for measuring ACh or ChE levels. Recently, however, Rubin (1960) proposed that speed of pupillary constriction may be taken as an indicator of the amount of the cholinergic mediator

(ACh) present. It should be recalled that synapses employing ACh as a transmitter are called cholinergic, while those employing norepinephrine are called adrenergic (Von Euler, 1959). The pupillary response provided a good measure of cholinergic mediator, because, as Rubin pointed out, "the magnitude of constriction is an increasing monotonic function of the amount of the cholinergic mediator liberated [p. 567]"; that is, faster constriction is indicative of greater amounts of ACh at cholinergic synapses. In like manner, Rubin suggested that speed of pupillary dilation be used as a measure of the amount of the adrenergic mediator (norepinephrine) present at the synapses, for again it seemed justified to assume a monotonic relationship (Rubin, 1964). The introduction of these measures has thus provided what seems to be simple and reliable measures of the transmitter substances and, consequently, predictors of learning.

If pupillary response is a predictor of learning in the human, then the application of this predictor might be carried one step further, into the area of personality. It has been theorized that an individual who learns or conditions rapidly tends to be highly sensitive to his environment, anxious, inhibited, compulsive, introspective, and often ill at ease. Conversely, the individual who conditions slowly would tend to be impulsive, irresponsible, unreliable, and insensitive to the environment and to the feelings of others. A considerable amount of research has been accumulated which supports the relationship between this dimension of personality and speed of

¹ Now at the University of Texas.

The author wishes to express his appreciation to Lee Sechrest for his assistance in the design of the research and to E. Femme for her assistance throughout the project.

conditioning (Eysenck, 1960; Franks, 1956; Johns & Quay, 1962; Quay & Hunt, 1965). Eysenck (1947) has referred to easily conditioned individuals as introverts, and to less easily conditioned individuals as extraverts; he has subsequently developed a scale to measure these characteristics (Eysenck, 1956).

The present study was an attempt to test the relationships between speed of neural transmission, conditionability, and personality. On the basis of earlier work with animals and the theoretical extrapolation to personality, it was predicted that subjects with inferred higher levels of the transmitter substances, as measured by the pupillary responses, would (a) be more aware of an environmental contingency, (b) evidence superior performance in verbal conditioning, and (c) rate themselves and be rated by peers as generally more introverted in personality.

METHOD

The subjects were 49 women in an introductory psychology class at Northwestern University. The first part of the experiment consisted of a verbal conditioning procedure similar to the one reported by Taffel (1955). Each subject was given 80 3×5 cards. On each card was typed a different past-tense verb. Below the verb were the pronouns I, we, you, he, she, and they. The pronouns appeared in a different order on each card. The subject's task was to make up a sentence using the verb and the pronoun of her choice. The experimenter sat behind a screen which separated him from the subject and recorded the pronouns used. The experimenter did not reinforce any of the subject's responses on trials 1 through 20. For trials 21 through 80, the experimenter responded with the word "good" in a flat, unemotional tone at the end of any sentence starting with the pronouns I or we. After completing the conditioning procedure, each subject was questioned concerning her awareness of the reinforcement contingency, according to a modified form of the inquiry schedule developed by DeNike and Spielberger (1963). In addition to assessing awareness, two items of the inquiry were used to determine the subject's behavioral intentions; that is, if the subject was aware of the reinforcement contingency, did the knowledge have a conscious effect on her behavior.

Following the verbal conditioning procedure, photographic measurements were taken of pupillary dilation and constriction. During the photographic procedure, the subject sat in a chair with her head held firmly in a Bausch and Lomb chin rest and forehead brace. The photographs of the eyes were taken with an Asahi Pentax single-lens reflex camera which was fitted with an extension bellows and an f4.5, 105 millimeter lens. The lens was 16 inches from the subject's eyes. The adapting light used to constrict the pupils was placed 9 inches from the subject's eyes, at a point 5 inches below the plane between the subject's eyes and the camera lens. This

light source consisted of a 3.5-inch square of white frosted glass which was illuminated from behind by a 100-watt white bulb. The lens of this light source was tilted so that the light was aimed directly at the subject's eyes. Eight and one-half inches to either side of the camera lens, and 14 inches from the subject's eyes were the infrared light sources, which were 2.75 inch squares. The illumination for these lights was provided by 200-watt bulbs which were housed in 10-inch photoflood reflectors. The light exposed to the subject was filtered through a Kodak Wratten filter No. 87. Consequently, the only light visible to the subject during the photographing of the dilating pupils was a dull red glow on either side of her field of vision. The intensity of the visible light was less than .5 footcandle.

All subjects were given the following explanation of the apparatus and procedure:

This apparatus will be used for taking photographic measurements of your eye. During the time the photographs are being taken, I will want you to keep your chin firmly in the chin rest and your forehead against the metal brace. While in this position, look straight forward at the lens of the camera. If you cannot see the lens because it is too dark, look straight forward at the point at which you think the lens should be. A series of photographs will be taken with this white light on and a series of photographs will be taken with this light off. When the light is off, you may notice a very dim red glow on each side of the side panels. Ignore these and continue to look straight ahead.

After the subject had her head positioned in the head rest, the room light was extinguished. Ten seconds later the adapting light and the infrared lights were turned on for 15 seconds to maximally constrict the pupil. At the end of this interval, the adapting light was turned off. Photographs of the dilating pupils were taken after intervals of 5, 10, 15, 30, 45, and 60 seconds. Prior to photographing the constricting pupils, the pupils were dark adapted for 1 minute. After 1 minute of complete darkness the adapting light and the infrared lights were turned on. The constricting pupils were photographed after intervals of 1, 2, 3, 4, and 5 seconds with a different trial being used for each measurement. The time intervals were controlled by a Hunter interval timer. Pupillary size after the various intervals of time was determined by projecting the developed negatives on a screen and measuring the diameter of the pupil with a millimeter ruler. The negatives were projected from a distance of 17 feet with a 5-inch lens. Since the photographs of the pupil carried only an identification number and the measurements were recorded on numbered data sheets, the experimenter did not know the identity of the subject who was being measured.

After completion of the photographic procedures, the subjects filled out a questionnaire which contained the Extraversion (E) and Neuroticism (N) scales from the Maudsley Personal Inventory, and the A, R, and MA scales from the MMPI.

Before leaving the laboratory, each subject was given three peer-rating forms and three envelopes. She was instructed to give one form and envelope to each of the three girls whom she felt knew her best and ask them to rate her on the various items. The form consisted of 16 descriptive statements about personality characteristics. All of the items were worded in the positive direction. Eight items described introverted characteristics and eight extraverted characteristics. Within each set of eight items, half were socially desirable and half less desirable, though none were completely undesirable. Rating of a subject on this form was done by placing an X on a 9-inch graphic scale beneath each item. The scale had end points labeled "Very Descriptive" and "Not Descriptive." All rating forms were returned to the experimenter in sealed, self-addressed envelopes. The scoring key was constructed so that extraverted descriptions received higher scores. The final score on the peer rating form was the sum of the ratings given by all three judges.

RESULTS

Speed of Pupillary Constriction

The mean and *SD* of the proportion of constriction after each interval are presented in Table 1. It should be pointed out that 100% constriction was not reported in this table because of what might be called a "bounce effect"; this is, some subjects reached maximal constriction in 3 or 4 seconds and then dilated slightly, which caused a drop in the constriction curve for these subjects and lowered the overall mean proportion of constriction after 4 and 5 seconds. Because of this bounce, the measurement after 4 and 5 seconds for some subjects was not an accurate measure of the speed of constriction, and therefore the measures taken after 1, 2, and 3 seconds were judged to be most accurate for comparing rates of constriction.

Extreme subjects (fast and slow) were identified as those whose proportion of constriction after at least two of the three intervals was plus or minus one *SD* from the mean of the distribution for each respective interval. Intermediate subjects were identified as those whose

proportion of constriction was within plus or minus one *SD* of the mean of each respective distribution. Eight subjects were identified as fast constrictors, 17 as intermediate speed constrictors, and eight as slow constrictors.

Speed of pupillary constriction and awareness. On the basis of their responses to the awareness inquiry, the subjects were classified as either aware or not aware of the reinforcement contingency. (It should be noted that there was 98% agreement between two independent judges with regard to classifying the subjects on awareness or behavioral intentions.) To test the hypotheses relating increased awareness and high ACh levels, the proportions of aware and not aware subjects within the fast and slow constricting groups were compared. A Fisher test of exact probability indicated that there was a significant (.05) relationship between fast constriction and awareness. Of the fast constrictors, 62.5% were aware, while 12.5% of the slow constrictors were aware.

Speed of pupillary constriction and verbal conditioning. Before discussing the relationship between speed of constriction and performance curves in verbal conditioning, some comment should be made concerning the subjects' conscious attitude toward performance on this task. When the postconditioning inquiries were scored for behavioral intentions, it was found that some of the aware subjects had consciously fought against giving conditioned responses. This tendency on the part of some subjects and the effect it had on the data has been noted and reported earlier (Farber, 1963; Holmes, 1966). Since the use of subjects who reported consciously resisting the giving of conditioned responses would distort any relationship between performance and speed of constriction, in Figure 1 where performance curves for fast, intermediate, and slow constrictors are plotted, only those subjects who did not resist conditioning were used.

Analysis of variance for repeated measures (Edwards, 1960) indicated that there were significant differences between the groups ($F = 53.54$, $df = 2,23$, $p < .01$) across blocks of trials ($F = 5.11$, $df = 3,69$, $p < .01$), and the Groups \times Trials interaction was significant ($F = 3.57$, $df = 6,69$, $p < .01$). The correlations between the total number of conditioned responses given on trials 21-80 and the proportion of constriction after intervals of 1, 2, and 3 seconds were .33 ($p < .05$), .44 ($p < .005$), and .52 ($p < .005$), respectively. All of these results clearly indicated that there was a significant relationship between fast pupillary constriction, which was indicative

TABLE 1
PROPORTION OF CONSTRICTION AFTER
INTERVALS OF TIME

Sec.	<i>M</i>	<i>SD</i>
1.0	63.69	10.15
2.0	82.73	8.07
3.0	92.76	5.35
4.0	98.10	2.57
5.0	98.10	4.78

Note.—*N* = 49.

of an inferred high ACh level, and performance on a verbal conditioning task. (It should be noted that since awareness is related to verbal conditioning performance—Farber, 1963; Holmes, 1966; Spielberger, 1965—the findings relating the personality response to awareness and to verbal conditioning performance are not completely independent. They were presented separately in this paper because while related, they represented different types of functioning and behavior which have been of interest to psychologists.)

Speed of pupillary constriction and personality measures. When the fast and slow constrictors were compared over the personality measures, the slow constrictors were found to be more extraverted than fast constrictors, as measured both by the E scale ($t = 3.09$, $df = 14$, $p < .01$) and by the peer-rating form ($t = 1.78$, $df = 14$, $p < .10$). Fast and slow constrictors did not differ significantly on the measures of general maladjustment, that is, the A, MA, and N scales ($t = .41$, .83, .43, respectively; $df = 14$).

In summary, fast pupillary constriction which is indicative of high levels of ACh was significantly related (a) to the development of awareness of a reinforcement contingency, (b) to superior performance on a verbal conditioning task, and (c) to the personality characteristics of introversion.

Speed of Pupillary Dilation

The mean and SD of the proportion of dilation after each of six intervals are presented in

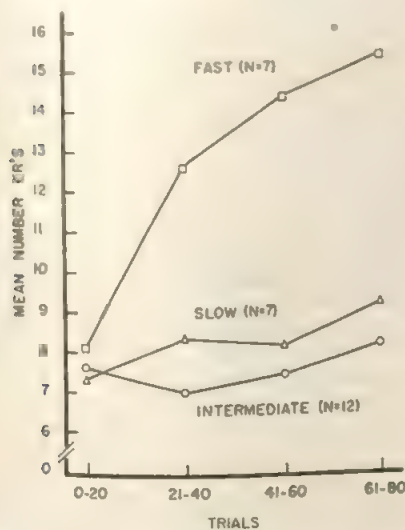


FIG. 1. Verbal conditioning performance curves for subjects with fast, intermediate, and slow pupillary constriction.

TABLE 2
PROPORTION OF DILATION AFTER
SIX INTERVALS OF TRIALS

Sec.	M	SD
5	77.55	8.74
10	81.56	6.99
15	92.45	5.85
30	95.44	4.43
45	97.89	3.97
60	98.86	1.81

Note.— $N = 49$.

Table 2. Because there was only limited variability in the subjects' responses after the intervals of 30, 45, and 60 seconds, extreme subjects were selected on the basis of their dilation after 5, 10, and 15 seconds. Fast ($N = 8$), intermediate ($N = 19$), and slow ($N = 6$) dilators were then selected in the same way as extreme constrictors.

Speed of pupillary dilation and awareness. A Fisher test of exact probabilities comparing the frequencies of aware and not aware subjects within the fast and slow dilating groups indicated that the distribution was well within the limits of chance occurrence.

Speed of pupillary dilation and verbal conditioning. When comparing the performance curves of the fast, intermediate, and slow dilators, the data from aware subjects who consciously resisted conditioning were not used. Analysis of variance indicated that there was a significant increase in conditioned responses over trials ($F = 5.56$, $df = 3, 63$, $p < .01$), but that there were no significant differences between the groups ($F = .06$, $df = 2, 21$).

Speed of pupillary dilation and personality measures. When the fast and slow dilators were compared over the personality measures, the fast dilators evidenced significantly higher E-scale scores ($t = 2.70$, $df = 12$, $p < .05$), but the groups did not differ on the peer-rating form measure of extraversion ($t = .51$, $df = 12$). As was the case with subjects differing in speed of constriction, the fast and slow dilators did not differ on the A, N, or MA scales ($t = 1.4$, 1.62, 1.54, respectively; $df = 14$).

DISCUSSION

The results of the present study clearly related the speed of pupillary constriction to awareness of an environmental contingency, susceptibility to verbal conditioning, and to personality in terms of the introversion-extraversion dimension. Since more rapid constriction is supposedly indicative of greater amounts of ACh at the

cholinergic synapses, the results of this study were in agreement with earlier findings on animals where subjects with an inferred higher level of ACh evidenced better performance. The results of the present study, however, extended the findings to the area of personality, and offered support for the hypothesis relating high levels of ACh to environmental awareness, rapid conditioning, and consequently to introverted personality characteristics. That is, subjects with inferred high ACh levels who are more aware and more easily conditioned by environmental contingencies would be more likely to learn, or introject, the rules, restraints, and anxieties of the environment than would their low ACh counterparts. The differences in neural conductivity, and the concomitant differences in reactions and conditioning, would therefore play a major role in their personality characteristics with regard to the dimension of introversion and extraversion.

Some comments might be made in regard to the personality differences which the present author found to be associated with fast and slow constriction. Rubin (1964) recently reported differences in constriction between normals and neurotics. The points on his Figure 2 (p. 564) indicated that after each interval the neurotics had evidenced more constriction than had the normals; that is, the neurotics were faster constrictors. Though consistent, the differences were not significant. When the diagnoses of the neurotic subjects were checked (p. 563), it was found that all of these subjects could be classified as suffering from *introverted* disorders. In the present study fast constrictors were significantly more introverted. The independent-dependent variables have been reversed in the two studies, but when Rubin's neurotics are seen as introverted, the results of the two studies are consistent. The fact that Rubin's findings were not significant could be attributed to the fact that his introverted neurotics were compared to an unselected sample of normals, a sample which, it is fairly certain, would have included introverted normals who would thus have decreased the differences between the groups.

The results using speed of dilation, the measure of the adrenergic mediator, were for the most part inconsistent and nonsignificant. This probably stemmed from the fact that "only the cholinergic transmitters have been proved to be transmitters within the brain," while the adrenergic transmitters may be limited to the autonomic system (Morgan, 1965, p. 554); that is, adrenergic synapses may not have played a role in the responses measured in this study.

The consistency of the findings within this

study, as well as the agreement between the results of this project and earlier research on animals, was very encouraging. It seems clear that if the pupillary response is in fact a function of ACh, the level of this transmitter substance is an important determinant of psychological functioning.

REFERENCES

- DeNIKE, L., & SPIELBERGER, C. Induced mediating states in verbal conditioning. *Journal of Verbal Learning and Verbal Behavior*, 1963, 1, 339-345.
- ECCLES, J. *The physiology of nerve cells*. Baltimore: Johns Hopkins Press, 1957.
- EDWARDS, A. *Experimental design in psychological research*. New York: Holt, Rinehart & Winston, 1960.
- EULER, U. VON. Autonomic neuroeffector transmission. In J. Field, H. Magoun, & V. Hall (Eds.), *Handbook of physiology*. Vol. 1. Washington, D. C.: American Physiological Society, 1959. Pp. 215-237.
- EYSENCK, H. *Dimensions of personality*. London: Kegan, Paul, 1947.
- EYSENCK, H. The questionnaire measurement of neuroticism and extraversion. *Revista di psicologia*, 1956, 50, 113-140.
- EYSENCK, H. *Experiments in personality*. London: Routledge & Kegan Paul, 1960.
- FARBER, I. The things people say to themselves. *American Psychologist*, 1963, 18, 185-197.
- FRANKS, C. Conditioning and personality: A study of normal and neurotic subjects. *Journal of Abnormal and Social Psychology*, 1956, 52, 143-150.
- HOLMES, D. Awareness, intentions, and verbal conditioning performance. Paper read at Midwestern Psychological Association, Chicago, May 1966.
- JOHNS, J., & QUAY, H. The effect of social reward on verbal conditioning in psychopathic neurotic military offenders. *Journal of Consulting Psychology*, 1962, 26, 217-220.
- MORGAN, C. *Physiological psychology*. New York: McGraw-Hill, 1965.
- QUAY, H. C., & HUNT, W. A. Psychopathy, neuroticism, and verbal conditioning: A replication and extension. *Journal of Consulting Psychology*, 1965, 29, 283.
- ROSENZWEIG, M., KRECH, D., & BENNETT, E. Brain chemistry and adaptive behavior. In H. Harlow and C. Woolsey (Eds.), *Biological and biochemical bases of behavior*. Madison: University of Wisconsin Press, 1958. Pp. 367-400.
- ROSENZWEIG, M., KRECH, D., & BENNETT, E. Heredity, environment, brain chemistry, and learning. In R. Patton (Ed.), *Current trends in psychological theory: A bicentennial program*. Pittsburgh: University of Pittsburgh Press, 1961. Pp. 87-110.
- RUBIN, L. Pupillary reactivity as a measure of adrenergic-cholinergic mechanisms in the study of psychotic behavior. *Journal of Nervous and Mental Disease*, 1960, 130, 386-400.

- RUBIN, L. Autonomic dysfunction as a concomitant of neurotic behavior. *Journal of Nervous and Mental Disease*, 1964, 138, 558-574.
- RUSSELL, R., WATSON, R., & FRANKENHAEUSER, M. Effects of chronic reductions in brain cholinesterase activity on acquisition and extinction of a conditioned avoidance response. *Scandinavian Journal of Psychology*, 1961, 2, 21-29.

- SPIELBERGER, C. Theoretical and epistemological issues in verbal conditioning. In S. Rosenberg (Ed.), *Directions in psycholinguistics*. New York: Macmillan, 1965. Pp. 149-200.
- TAFFEL, C. Anxiety and the conditioning of verbal behavior. *Journal of Abnormal and Social Psychology*, 1955, 51, 496-501.

(Received October 6, 1965)

Journal of Personality and Social Psychology
1967, Vol. 5, No. 1, 103-107

AROUSAL CORRELATES OF TASK ROLE AND GROUP SETTING¹

DAVID SHAPIRO

AND

P. HERBERT LEIDERMAN²

Harvard Medical School

Stanford University School of Medicine

Success and failure task roles were studied in different group settings with each individual's overt activity held constant. Significant effects were found for electrodermal level, a measure of behavioral arousal, but not for heart rate. There was greater arousal, as shown by slower rate of physiological habituation, in failure compared to success and in paired compared to single roles. These 2 variables also had an effect on role perceptions. It was suggested that persistent arousal was more likely in roles characterized by uncertainty or ambiguity.

One recent development in research on social interaction is represented by studies in which both physiological and group behavioral measures are obtained (Leiderman & Shapiro, 1964a). To the extent that measures of autonomic effector activity can be considered indexes of drive or arousal, they add an important source of information. A combined analysis of behavioral outcomes and variations in arousal state may be brought to bear on a number of theoretical issues in social psychology.

An old problem which has recently been revived concerns the effects of the presence of other persons on individual performance. To account for these effects, on the basis of indirect

evidence from animal studies, Zajonc (1965) hypothesized that arousal is increased by the presence of others. Using electrodermal level as a physiological index of alertness or behavioral arousal in human subjects (Leiderman & Shapiro, 1964b), we have found confirmation for this hypothesis in a task situation. Arousal levels were higher when individuals were performing a group task than when they were engaged in the same task alone (Shapiro, Leiderman, & Morningstar, 1964). Unpublished data from this study, however, suggested that the mere presence of two other persons during a preexperimental rest period did not in itself lead to higher levels of electrodermal activity compared to waiting alone.

Other small-group studies indicate that a number of physiological indexes of arousal may be either increased or decreased depending on such variables as the nature of the group task and its meaning to the individual (Back & Bogdonoff, 1964), the mutual attitudes of the interacting parties (Kaplan, Burch, & Bloom, 1964), the degree to which an individual succeeds or fails in a group task (Leiderman & Shapiro, 1963), and the composition of roles in a group (Shapiro & Leiderman, 1964a).

Granted that variations in physiological arousal can be socially determined, a question arises as to the dependence of these effects on the actual

¹The research was supported by Office of Naval Research Contract Nonr-1866(43), Group Psychology Branch, and by United States Public Health Service Research Grant MH-08853-01 and Career Development Award K3-MH-20,476-02, National Institute of Mental Health.

Use of the facilities at the Harvard University Computing Center was made possible through National Science Foundation Grant NSF-GP-683.

²We thank Bernard Tursky, instrumentation engineer, for the design of experimental apparatus and many helpful suggestions on all phases of the research. We are grateful to Andrew Crider and Fredric Levine for a critical review of the manuscript. Technical and research assistants were Anne Crider, Kevin Deasy, Arlene Safft, and David Wolfe.

individual behaviors induced by the social situation. To eliminate the influence of behavioral activity, in the present experiment all subjects performed the same behavioral responses during a group task. By manipulating the reward value of these responses, two different task roles were derived experimentally. These roles were embedded in two different group settings which were composed of different combinations of the two roles. Using electrodermal level as an index of arousal, the "pure" arousal effects of different task roles and of different social settings were assessed. Data were also gathered on individual appraisals of their performance following the task in order to study the relationship between role perception and the physiological changes under the controlled behavioral conditions.

PROCEDURE

A guessing game was used as the experimental task. It required subjects in three-person groups to guess, one at a time in random order, which one of six colors would be next on the experimenter's list. If a guess was correct, they were told they would hear a tone and if incorrect, a buzzer. They were also told that the object of the game was to guess as many correct as possible and that their performance as a group would be compared to other groups of subjects selected from the same population. The three persons in each group sat around a table in face-to-face contact. They could hear each other's guesses and the outcome (tone or buzzer) of each trial, but were not permitted to discuss their decisions.

Unknown to the subjects there was no list of colors. Success and failure were manipulated by the experimenter: In a successful role, a subject's guesses were reinforced with a tone on 5/6 of the trials at random and punished with a buzzer on the remaining trials; in a failure role, a subject's guesses were punished on 5/6 of the trials and reinforced on the remaining trials. The task period consisted of 90 trials evenly spaced at 2 a minute, 30 per subject, or 45 minutes in all. Two different group settings were also contrived. In the first, two subjects were successful and the other a failure; in the second, two were failures and the other a success. Eleven groups were run in the first and nine in the second setting, giving 22 *paired-success*, 11 *single-failure*, 18 *paired-failure*, and 9 *single-success* roles. Each subject participated in one role condition only, making a total of 60 subjects. The subjects were student nurses, 18-21 years old, paid \$1.50 per hour.

The following instructions were used:

Now we want you to take part in a GROUP PROBLEM-SOLVING task. This is how it works. In the other room I have a long list of colors written down, and you have to guess each color on my list and try to get as many right as possible.

Each time one of you sees the word START that person mentions a color choice, then all three

of you press the appropriate button of that color. All three of you must press at the same time. The color chosen will appear in front of you. The colors are white, amber, yellow, red, green, and blue. If you chose correctly, you will hear a sound like this [tone]. If you were incorrect you will hear a sound like this [buzz]. Whether you are correct or incorrect, I will go on to the next color on my list and you have to decide what it is.

Now I will give you a few practice trials. Watch for START and respond when you see it. [three practice trials] The time between responses is to be used to sit without talking and think out a solution to the problem.

Remember that the person who sees the START signal must name the color on my list and try to get as many right as possible. Your performance as a group in this problem solving task will be compared to other groups.

There will now be a rest period. At the end of this period I will sound a long tone. After this tone watch for the first START signal.

After the instructions were read, there was a 15-minute rest period to establish physiological base lines. The 45-minute total experimental session which followed was divided into four equal time periods for purposes of analysis.

The measure of physiological arousal was skin potential (SP) level which is sympathetically innervated electrodermal activity (Leiderman & Shapiro, 1964b). It was recorded as the direct-current potential from an active area on the thenar eminence of the left palm with an inactive area on the dorsal aspect of the left forearm approximately 20 centimeters from the palm as the reference. Nonpolarizing silver-silver chloride sponge electrodes were used (O'Connell & Tursky, 1960). The continuous SP level was treated as a slowly changing function and sampled at 1-minute intervals. Difference scores ($\Delta \bar{X}$ SP) between the mean level of the rest period and the mean level for each of the four experimental periods were analyzed (Shapiro & Leiderman, 1964b).

Heart rate was obtained as a second physiological measure via electrocardiograph electrodes on the upper arms. The number of beats per 12-second block at each minute sampling point was counted to estimate heart rate for each minute. Difference scores were used as in the analysis of SP level.

The laboratory was a sound-attenuated, light-and-temperature controlled, 9 × 13-foot room equipped with a one-way-vision mirror. The six physiological measures (two per subject) were recorded continuously during each session on an Offner Type R dynagraph.

Following the session, a brief questionnaire was administered in which subjects were asked to evaluate their performance using three simple rating scales:

1. How well do you think you did?
2. How well do you think you did compared to how well you should have done?

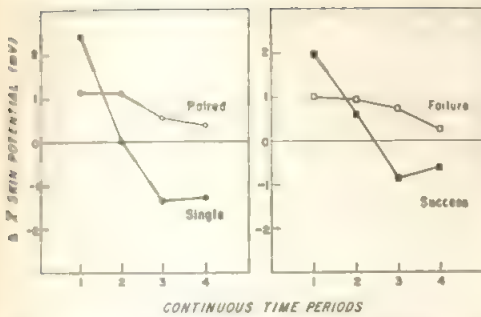


FIG. 1. Interaction between two role variables and time.

3. How well do you think you did compared to how well you expect other student nurses to do?

RESULTS

The effects of the three independent variables were tested by the analysis of variance for repeated measures (Edwards, 1950): role reinforcement (success, failure), role pairing (paired, single), time (four periods). For SP level the main effect for time was significant at the .01 level, and the interaction of Time \times Pairing was significant at the .05 level. The interaction of Time \times Reinforcement approached significance. In order to obtain a second estimate of the significance of this latter interaction, the analysis was recomputed omitting the fourth time period. This was done because, as Figure 1 suggests, the interaction effects were probably maximum by the third time period. The curves reached stable values in this period with the likelihood that subsequent variation resulted from other factors such as boredom with the task that were not intrinsic to the experimental role conditions. The second analysis (Table 1) confirmed the first and showed in addition that the interaction of Role Reinforcement \times Time attained significance at the .05 level.

TABLE 1
ANALYSIS OF VARIANCE—SKIN POTENTIAL LEVEL
THREE EXPERIMENTAL TIME PERIODS

Source	SS	MS	F
Reinforcement (R)	5.09	5.09	0.06
Pairing (P)	12.06	12.06	0.15
R \times P	36.34	36.34	0.44
Error (a)	4575.73	81.71	
Time (T)	79.66	39.83	5.38**
R \times T	49.46	24.73	3.34*
P \times T	72.85	36.42	4.92**
R \times P \times T	15.56	7.78	1.05
Error (b)	828.47	7.40	

* $p < .05$, $df = 2/112$.

** $p < .01$, $df = 2/112$.

TABLE 2
SUBJECTIVE RESPONSES

Task role	How well do you think you did compared to how well you feel you should have done?				
	Very well	Well	Average	Poor	Very poor
Single success	1	6	1	1	0
Paired success	3	5	12	2	0
Single failure	0	2	2	1	6
Paired failure	0	0	9	9	0

$$\chi^2 = 57.70, df = 12, p < .001$$

Success	4	11	13	3	0
Failure	0	2	11	10	6

$$\chi^2 = 21.28, df = 4, p < .001$$

Paired	3	5	21	11	0
Single	1	8	3	2	6

$$\chi^2 = 23.47, df = 4, p < .001$$

A subject-by-subject analysis of the number of individuals in the different roles who habituated physiologically from the first to third time periods supported the above statistical analyses and suggested that the trends graphically depicted in Figure 1 were typical of individual subjects.

The major results of the experimentally derived roles, thus, are shown not in overall level differences but in differential habituation of SP level over time as a function of role reinforcement and role pairing.

Comparable analyses of the heart-rate data did not yield any findings of statistical significance, and these results will not be presented.

On the questionnaire the mean response to each of the three items was approximately the same within each role. Data on the second item differentiated the roles the best (Table 2). The results show that subjects were differentially aware of the reward value of their activity, and their perceptions varied significantly with the group setting. Paired subjects were less extreme in their reports than subjects in single roles. Only 15% of the latter said their performance was "average" as compared to 52.5% of the paired subjects ($p < .05$). One was more likely to regard his performance as average when another subject shared the same role.

DISCUSSION

The results confirm previous findings that task roles are significantly related to changes in

electrodermal activity (SP level), but independent of heart rate (Shapiro & Leiderman, 1964a). The effects are shown in varying degrees of physiological habituation to the situation, defined by a decline in electrodermal activity. The physiological differentiation of role observed in the present study and in previous research cannot be accounted for by differences in overt behavior, which was the same for all subjects in the present experiment. Rather, the differences must be attributed to success and failure per se and to the group context in which they are embedded.

Assuming SP level to be an index of behavioral arousal (Leiderman & Shapiro, 1964b), we infer that task roles lead to states of arousal which habituate at varying rates. Failure roles show less habituation than success roles. This agrees with Morningsstar's (1963) data in a comparable individual task. The second finding, that paired roles show less habituation than single roles, is also consistent with previous research (Shapiro & Leiderman, 1964a).

According to Spence (1956), impaired physiological habituation is a consequent of anxiety. This relationship is supported by empirical research linking differential physiological habituation to discrete stimuli with clinical assessments of anxiety (Lader & Wing, 1964). We can therefore infer a greater degree of anxiety in failure as opposed to success and paired as opposed to single roles.

While the greater arousal, that is, impaired habituation, in failure roles is easily understood, some interpretation is demanded to account for the greater arousal in paired as opposed to single roles. Postexperimental questionnaires showed that single-success subjects evaluated their performance as better than did paired-success subjects. By the same token, single-failure subjects saw their performance as poorer than paired-failure subjects. The objective degree of success and failure was identical for single and paired roles, of course. The tendency for paired subjects to use less extreme categories to rate their performance is indicative of some degree of uncertainty as to their relative group standing. Having little basis on which to judge their performance in the experimental task, subjects must have compared their roles to others in the group. Paired roles were embedded in a group setting in which one cohort performed similarly, and the other dissimilarly, making the comparison somewhat ambiguous. On the other hand, single-success or single-failure subjects were clearly dissimilar from their paired cohorts, accounting for their greater certainty in evaluating relative performance. It is likely that the ambiguity or uncertainty consequent upon a paired role is a

source of the more persistent arousal in these subjects.

Processes of social comparison have an effect on self-evaluations and the interpretation of emotional states (Festinger, 1954; Schachter, 1964). The present experiment suggests that the social setting influences role perception as well as the course of habituation of an autonomic effector system. Apparently these processes of comparison occurred in an experimental situation providing little opportunity for interaction and communication among the participants.

Finally, Zajonc (1965) has hypothesized that the mere presence of others increases arousal. The present experiment has shown that, given the same pattern of individual task behavior, the effects of the presence of others on arousal varies with the group setting, depending on the functional interrelationships of the participants in addition to the reinforcing properties of the situation. It is suggested that where the presence of others tends to decrease uncertainty or ambiguity about one's role in a situation, level of arousal may be reduced or its rate of habituation accelerated. This might explain why being with friends is less arousing than being with strangers (Back & Bogdonoff, 1964).

REFERENCES

- BACK, K. W., & BOGDONOFF, M. D. Plasma lipid responses to leadership, conformity, and deviation. In P. H. Leiderman & D. Shapiro (Eds.), *Psychobiological approaches to social behavior*. Stanford: Stanford University Press, 1964. Pp. 24-42.
- EDWARDS, A. I. *Experimental design in psychological research*. New York: Holt, Rinehart & Winston, 1950.
- FESTINGER, L. A theory of social comparison processes. *Human Relations*, 1954, 7, 117-140.
- KAPLAN, H. B., BURCH, N. R., & BLOOM, S. W. Physiological covariation and sociometric relationships in small peer groups. In P. H. Leiderman & D. Shapiro (Eds.), *Psychobiological approaches to social behavior*. Stanford: Stanford University Press, 1964. Pp. 92-109.
- LADER, M. H., & WING, L. Habituation of the psychogalvanic reflex in patients with anxiety states and in normal subjects. *Journal of Neurology, Neurosurgery, and Psychiatry*, 1964, 27, 210-218.
- LEIDERMAN, P. H., & SHAPIRO, D. A physiological and behavioral approach to the study of group interaction. *Psychomatic Medicine*, 1963, 25, 146-157.
- LEIDERMAN, P. H., & SHAPIRO, D. (Eds.) *Psychobiological approaches to social behavior*. Stanford: Stanford University Press, 1964. (a)
- LEIDERMAN, P. H., & SHAPIRO, D. Studies on the galvanic skin potential level: Some behavioral correlates. *Journal of Psychosomatic Research*, 1964, 7, 277-281. (b)
- MORNINGSSTAR, M. A study of some relationships between reinforcement and physiological responses.

- Unpublished master's thesis, Northeastern University, 1963.
- O'CONNELL, D. N., & TURSKY, B. Silver-silver chloride sponge electrodes for skin potential recording. *American Journal of Psychology*, 1960, 73, 302.
- SCHACHTER, S. The interaction of cognitive and physiological determinants of emotional state. In P. H. Leiderman & D. Shapiro (Eds.), *Psychobiological approaches to social behavior*. Stanford: Stanford University Press, 1964. Pp. 138-173.
- SHAPIRO, D., & LEIDERMAN, P. H. Acts and activation: A psychophysiological study of social interaction. In P. H. Leiderman & D. Shapiro (Eds.), *Psychobiological approaches to social behavior*. Stanford: Stanford University Press, 1964. Pp. 110-126. (a)
- SHAPIRO, D., & LEIDERMAN, P. H. Studies on the galvanic skin potential level: Some statistical properties. *Journal of Psychosomatic Research*, 1964, 7, 269-275. (b)
- SHAPIRO, D., LEIDERMAN, P. H., & MORNINGSTAR, M. Social isolation and social interaction: A behavioral and physiological comparison. In J. Wortis (Ed.), *Recent advances in biological psychiatry*. Vol. 6. New York: Plenum Press, 1964. Pp. 129-138.
- SPENCE, K. *Behavior theory and conditioning*. New Haven: Yale University Press, 1956.
- ZAJONC, R. B. Social facilitation. *Science*, 1965, 149, 269-274.

(Received December 30, 1965)

Journal of Personality and Social Psychology
1967, Vol. 5, No. 1, 104-110

WORD MEANINGS AND SELF-DESCRIPTIONS: A REPLICATION AND EXTENSION

JOHN C. LOEHLIN¹

University of Texas

An earlier study comparing the consistency with which college-student Ss assign meanings to adjectives and the consistency with which they describe themselves using the same adjectives was repeated and extended to pairs and triads of loosely synonymous adjectives. It was again concluded that self-ratings using adjectives provide weak tools for inferring self-concepts in such a population, since (a) the individuality in self-description which is consistent across measurement methods is not great; and (b) at least some of this consistency results from individual differences in the meanings attributed to the concepts used. Using multiple synonyms to define the meanings of concepts had, on the whole, little effect on the results.

In an earlier study (Loehlin, 1961) the writer examined the consistency among college-student subjects in the meanings they attributed to personality-descriptive words, and compared this to the consistency among the same subjects in the use of the same words in self-ratings. The major conclusion of that study was that the subjects differed fully as much in the meanings they attributed to the words as they did in the use of the words to describe themselves, when both were equivalently measured. Such a result suggests that considerable ambiguity is likely to result if self-ratings are used to make inferences about self-concepts, since it will be uncertain whether subjects are differing primarily in the meanings they attribute to the words they are

using, or whether they genuinely view themselves differently (cf. Kelly, 1955).

One purpose of the present study was to attempt replication of the earlier one. A second purpose was to explore one factor possibly bearing on the previous results. The words used in the earlier study were single adjectives descriptive of personality—words like *expressive*, *solemn*, *considerate*, *irresponsible*. Since a single adjective may cover a range of meanings, the differences among subjects could conceivably arise from the fact that each happened to select some particular meaning from the range and used it consistently in his ratings. If, instead of single words, two (or more) adjectives overlapping in meaning are used, it should be possible to restrict the range of interpretation to their common core of meaning. Thus, *expressive* may mean "self-revelatory" to one person and "eloquent" to another, but the combination *expressive, open, frank* should be less ambiguous.

¹The data for this study were collected while the writer was at the University of Nebraska. I wish to express my appreciation for the help of Evelyn Thoman in several aspects of the data gathering and processing.

A decrease of individual differences in concept meanings, if individual differences in self-description remain substantial, would improve prospects for using the concepts in making inferences about the self. A decrease in both cases, or in neither, would of course leave the situation much the same as before.

METHOD

The procedures of the present study followed closely those of the earlier one. Again college-student subjects in a classroom setting made various ratings using adjectival concepts. The experiment was carried out with three different sets of six concepts; each of these replications was with a group of subjects presumably comparable to those of the original study (students taking the same introductory personality course, in different semesters). A random third of the subjects on each occasion received questionnaires identical to corresponding ones from the original study (single adjectives), another third made similar ratings using adjective pairs, and the final third made ratings using adjective triads. The sets of adjectives used are listed in Table 1, in which the triad version is shown. The first adjective was used for the single version in each case, the first two for the pairs. The single adjectives are the ones used in the previous study, where they were obtained by random sampling from a pool representing Cattell's (1957) "personality sphere." The second and third adjectives were selected by the writer to overlap with and specify some meaning of the first. Loose, rather than exact, synonyms were sought in an effort to restrict the range of common meaning while not losing it altogether.

The instructions were the same as those used earlier and were identical for the three versions

except for referring in the latter two to a "phrase which might be used to describe a person," rather than a "word which might be used to describe a person."

As in the earlier study, three different approaches were used to define what the words meant to each subject and to obtain his self-description: (a) direct rating of similarities among adjective meanings and applicability of the adjectives to the self, (b) semantic differential ratings (Osgood, Suci, & Tanenbaum, 1957) of the adjectives and the self, and (c) rating of other persons well known to the subject, using the adjectives, and comparison of the self to these persons. The data from each of these methods were used to locate the adjectives and the self in a "semantic space." In such a space similarities in meaning among concepts are represented by distances: concepts similar in meaning are close together, concepts opposite in meaning are far apart (Osgood et al., 1957).

To permit comparison of the resulting semantic spaces across individuals and across methods, despite differing scale units and individual response sets, each space was described by rank ordering its distances. This was done separately for the 15 distances between all possible adjective pairs, and the 6 distances between adjectives and the self. Agreement among persons concerning the ordering of interadjective distances was taken as an index of agreement on adjective meanings. Agreement among persons concerning the ordering of adjective-self distances was taken as an index of agreement on self-descriptions. Extent of consistency among a group of rankings was assessed by the average Spearman rank-order correlation (ρ) among them. This can be obtained by averaging ρ 's directly, but is more conveniently calculated from Kendall's coefficient of concordance (W) and the relation $\bar{\rho} = (mW-1)/(m-1)$, where m is the number of rankings involved (Kendall, 1948).

The main conclusions of the present study are based on comparisons of the consistency across methods within subjects with that across both subjects and methods. The first of these reflects the consistency of individual subjects' concept meanings or self-descriptions. It can be obtained by computing W for each subject across the three methods, converting to $\bar{\rho}$, and averaging over subjects. The second reflects group consensus on concept meanings and self-descriptions. It is most conveniently obtained indirectly: Given the mean of ρ 's for all comparisons among rankings, and the means for two discrete subsets of these comparisons (within subjects across methods, and within methods across subjects), the mean for the remaining comparisons, those made across subjects and methods, is easily calculated. The mean ρ 's needed can be obtained with Kendall's W as described above. An example of the calculation is given in the previous article (Loehlin, 1961).

RESULTS

The data of principal interest are presented in Table 2. For each sample we can ask: (a) How consistently do subjects describe themselves, and

TABLE 1.

LISTS OF ADJECTIVES USED

Sample A

Complaining, sensitive, distressed
Dominant, assertive, masterful
Imaginative, fanciful, playful
Relaxed, easygoing, placid
Inconsiderate, rude, thoughtless
Expressive, open, frank

Sample B

Submissive, meek, compliant
Solemn, sober, serious
Expressive, open, frank
Considerate, courteous, thoughtful
Inflexible, stiff, rigid
Obstructive, uncooperative, difficult

Sample C

Inflexible, stiff, rigid
Dependent, clinging, reliant
Responsible, dependable, trustworthy
Tough, unsentimental, hardheaded
Persevering, determined, thorough
Irresponsible, neglectful, undependable

TABLE 2

CONSISTENCY ACROSS MEASUREMENT METHODS AMONG SELF-DESCRIPTIONS AND CONCEPT MEANINGS

Comparisons	Mean ρ among rankings of distances							
	Self				Concepts			
	1961	Single	Pair	Triad	1961	Single	Pair	Triad
Complaining, etc.								
Within Ss	.48	.40	.56	.54	.44	.47	.40	.40
Across Ss	.38	.28	.30	.33	.36	.42	.32	.30
Individuality	.10	.12	.26	.21	.08	.05	.08	.10
<i>N</i>	(19)	(23)	(23)	(17)				
Submissive, etc.								
Within Ss	.40	.26	.49	.48	.43	.30	.32	.36
Across Ss	.21	.19	.39	.34	.28	.24	.26	.28
Individuality	.19	.07	.10	.14	.15	.06	.06	.08
<i>N</i>	(17)	(15)	(19)	(16)				
Inflexible, etc.								
Within Ss	.50	.45	.47	.48	.60	.58	.43	.45
Across Ss	.40	.39	.43	.34	.34	.38	.38	.37
Individuality	.10 ¹	.06	.04	.14	.26	.20	.05	.08
<i>N</i>	(16)	(13)	(19)	(17)				

how consistently do they assign meanings to concepts? (b) How much of this consistency in each case is common to the group and how much is idiosyncratic? (c) Is the individuality in self-description enough greater than that in word meanings so that the latter can reasonably be used in interpreting the former? (d) And, finally, How are *a*, *b*, and *c* affected by rating pairs or triads instead of single adjectives?

Let us first consider the single adjectives: the first two columns on each side of the table. These represent the 1961 study and its replication in the present study. For each sample of adjectives, the first row of data represents the average individual consistency across methods, the second row represents consistency across methods and subjects (i.e., consensus), and the third row is the discrepancy between the two, representing consistent individual differences.

First, we may note that the replication does not duplicate the data from the original study exactly, but agrees well with it in general trend: a moderate degree of consistency across methods in the ratings by individual subjects, with a large proportion of this reflecting general consensus of ratings, and a distinctly smaller proportion reflecting consistent individuality. Finally, we may observe that individual distinctiveness in assigning meaning to concepts remains on the whole as high as individual distinctiveness in self-description, tending to confirm the main conclusion of the original study.

Next, let us examine the effects of rating pairs or triads rather than single adjectives. If we look

first at the right-hand side of the table, based on concept ratings, it is fairly clear that rating adjective pairs or triads has produced no striking tendency toward increased consistency of meaning, either consensual or idiosyncratic.

On the left-hand side of the table, things are less clear. There appears to be a moderate trend toward increased consistency as we go from single adjectives to triads, but since the obtained differences are in some instances no greater than the differences between the replications with single adjectives, such a trend must be inferred with considerable caution. It is, in any event, a mildly paradoxical result, since we appear to have here a sharper definition of the self-concept despite the fact that the terms used show no increased precision of meaning. It is of course conceivable that distance from the self is a more sensitive indicator of the meaning of a concept than is its distance from other concepts, but this is not particularly encouraging to the use of the concept in the definition of the self.

One final question concerns the relationship between the individual's description of himself and the individual's assignment of meaning to concepts. In the previous study, there appeared to be a tendency for the two to covary: for example, subjects who judged *solemn* to be relatively self-descriptive also judged it to be relatively close in meaning to *expressive* and *considerate* and relatively far from *submissive* and *obstructive*. For the first sample of adjectives in the present study, correlations were obtained between the rankings of adjective-self distances

and interadjective distances, for singles, pairs, and triads for each of the three measurement methods—a total of nine correlations representing each relationship. Since the *N*s were small (17–23), there was a good deal of fluctuation in these correlations, but 7 of the 30 relationships showed a reasonable degree of consistency, defined as at least 7 of the 9 correlations in the same direction, at least half of these over .20, and none of the negative cases this large. The consistent relationships thus obtained appeared to be of the same general character as those noted in the earlier study. Subjects describing themselves as relatively *complaining* tended to locate *complaining* nearer to *dominant*, *imaginative*, and *expressive*, and farther from *inconsiderate*; subjects describing themselves as relatively *dominant* located *dominant* nearer to *relaxed*; subjects describing themselves as relatively *inconsiderate* placed this concept nearer *expressive*; and those describing themselves as relatively *expressive* placed *expressive* closer to *imaginative*.

Thus, as in the earlier study, we again emerge with somewhat pessimistic conclusions about the meaningfulness of adjectival self-descriptions in defining the self-concept, at least for subjects and settings like ours, for two reasons: (a) The individuality in self-description which is consistent across measurement methods is not great; and (b) at least some of this appears to result from

individual differences in the meanings attributed to the concepts used.

The use of multiple synonyms in the present study did not materially increase the consensus among subjects in the meanings assigned to concepts and hence does not provide a ready solution to the problem. It is conceivable that other means of improving concept definition might do better, but it should be noted that the rather small amount of variation in self-description in this population implies that a considerable improvement would be necessary before differential inferences about self-concepts could be confidently made.

REFERENCES

- CATTELL, R. B. *Personality and motivation structure and measurement*. Yonkers-on-Hudson, N. Y.: World Book, 1957.
- KELLY, G. A. *The psychology of personal constructs*. Vol. 1. *A theory of personality*. New York: Norton, 1955.
- KENDALL, M. G. *Rank correlation methods*. London: Griffin, 1948.
- LOEHLIN, J. C. Word meanings and self-descriptions. *Journal of Abnormal and Social Psychology*, 1961, 62, 28–34.
- OSGOOD, C. E., SUCL, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: University of Illinois Press, 1957.

(Received February 4, 1966)

Journal of Personality and Social Psychology
1967, Vol. 5, No. 1, 110–114

COMPLEXITY OF ATTITUDE STRUCTURE AND PROCESSES OF CONFLICT REDUCTION¹

WILLIAM D. CRANO

AND

H. M. SCHRODER

Northwestern University

Princeton University

The manner in which 2 groups differentiated in terms of conceptual complexity of personality structure utilized available processes of conflict reduction in reaction to a counternorm message was studied. The low-complexity Ss (*N* = 17) tended to utilize the 6 available processes of message acceptance or rejection, assimilation-contrast, and source approbation or deprecation in an internally consistent fashion significantly more often ($p < .005$) than the high-complexity group (*N* = 14). Consistent with previous findings, the low-complexity Ss most often used the processes in a mutually exclusive manner. The high-complexity group, however, did not follow this trend ($p < .05$). Possible deficiencies of prediction in theories failing to consider these differing modes of conflict reduction were discussed.

This study investigated the possibility that the processes involved in attitude change vary as a

¹ This study was supported by Contract Nonr 1858(12), sponsored by the Group Psychology Branch, Office of Naval Research, H. M. Schroder, principal investigator.

function of the structure of the attitude. An individual's attitude is defined as a type of information-processing structure, and the principles involved in reacting to attitudinal discrepancy or conflict are expected to vary as the level or complexity of the structure varies. It was

hypothesized that the introduction of attitudinally discrepant information into a more integratively simple attitude structure (i.e., one having fewer differentiated dimensions of information, and fewer integrative linkages between the parts in the attitude) would generate processes different from those accompanying the introduction of the same discrepant information into a more integratively complex structure (having higher degrees of freedom in the organization of dimensional units involved in the attitude), when measures of direction and magnitude are constant for both structures. This hypothesis arises from a number of experiments which demonstrate that differences between the complexity of information-processing characteristics of persons (for a given range of stimuli) are accompanied by systematic changes in the behavioral consequences of similar antecedent conditions (Schroder, Driver, & Streufert, 1966).

Many aspects of this question have been studied, including various mechanisms of defense (Carlson, 1954; Cooper & Jahoda, 1947; Erikson, 1952, 1954; Harvey, Kelley, & Shapiro, 1957; Kelley & Woodruff, 1956) and modes of conflict reduction leading to attitude change (Abelson, 1959; Hunt & Schroder, 1958; Kelman, 1953, 1958; Osgood & Tannenbaum, 1955). Harvey et al. (1957) found the interrelations between three processes of conflict reduction to be essentially nonsignificant. This observation was consistent with the hypothesis of Kelley and Woodruff (1956) who proposed that:

The three reactions (compliance, distortion, and dissociation) constitute intersubstitutable ways of handling the pressure to change. . . . Because of their interchangeability as ways of handling the communication situation, we would expect them to occur in a negatively interdependent manner [p. 73].

This formulation would appear contrary to the dissonance-reduction hypothesis. Were one to follow the model of Festinger (1957), he would predict that resolution processes would tend to occur in an internally consistent manner—that is, all would act in the same direction, thereby facilitating the reduction of dissonance. For example, if an individual changes his attitude as a result of a communication and accepts its information, he would, if utilizing other processes, tend neither to deprecate the source of the information, nor distort the content of the message in a direction away from his established beliefs, for both responses would act to maintain conflict within the conceptual system. Further, if the reduction of dissonance is a primary objective, then the use of more than one resolution process

seems highly probable. If the discrepant information is accepted, then the utilization of source approbation and a distortion of the message toward the accepted position would further enhance the reduction of conflict.

Bieri and Blacker (1956) have approached this question in a slightly different manner by stating that as the number of alternative perceptions of a situation increase, so too do the number of alternative behaviors available to the individual. That is, the use of fewer or more resolution processes following the presentation of attitudinally discrepant material is predictable and is dependent upon the complexity of the perceptual structure of the individual.

The conflict-reduction processes to be considered in this report are acceptance-rejection, approbation-deprecation, and assimilation-contrast. Acceptance is defined as attitude change toward a position advocated by a counternorm message along a given continuum. The opposite pole of this reaction is referred to as rejection, through which the existing attitude is bolstered or changed in a direction opposite to that advanced by the attitudinally discrepant communication.

Assimilation refers to the perception of a message in such a way as to make it appear to be more in agreement with established or existing attitudes than, in fact, the message really is. It is also possible to distort the content of a message in such a way that it is perceived as more threatening to the belief system than is actually the case. This reaction has been called contrast (Sherif & Hovland, 1961) and represents a type of perceptual distortion, although it is a distortion in a direction opposite to that of assimilation.

If the source of a counternorm message is perceived by an individual to be, for example, non-authoritative or incompetent, then the person has utilized the evaluative process of source deprecation. If, however, the source is perceived as having more authority or expertise as a consequence of the message attributed to it, then the process of approbation has been used.

It is obvious that these processes may be combined in a number of different ways. If, however, an individual uses *only* those processes which tend to move him in a direction advocated by a given message (i.e., through some combination of assimilation, acceptance, and/or source approbation), or *only* those processes which tend to move him in the direction opposite to that advocated (i.e., through some combination of contrast, rejection, and/or source deprecation), then we have defined his patterning of processes as internally consistent. If,

however, processes of both types are utilized, then the combination has been labeled internally discordant.

HYPOTHESES

It should be made clear at this point that the study is concerned chiefly with the patterning of the processes which an individual utilizes, and not with the magnitude of these processes *per se*. The proposition that the number and consistency of these processes engaged in reacting to a discrepant message vary as some function of the level of conceptual complexity (Schroder et al., 1966; Suedfeld, 1964) was investigated. Specifically, it was hypothesized that an individual operating at a low level of complexity would attempt to resolve conflict through the utilization of processes in either a mutually exclusive or internally consistent manner. As the complexity of a person's information-processing structure increases, the individual will generate more degrees of freedom in dealing with diversity. Consequently, there will be a greater number of resolution processes available to him, and these will not necessarily be bound by the condition of internal consistency. That is, multidimensional and alternate integrative rules will operate at more complex levels.

METHOD

Schroder and Streufert's (1963) Sentence Completion Test was administered to 150 male undergraduates at Princeton University. Sentence completions to five stems which introduce conflict and uncertainty (e.g., When I am in doubt . . . ; When I am criticized . . . ; etc.) were scored on the basis of the level of information processing involved in the written responses, along a 7-point scale. A scale point of 1 represents a response which could be generated by a single fixed rule in which no alternative perspective appeared. A scale point of 5 indicates a response which contains alternate perspectives of the event and a consideration of their relationship. Scale points of 6 and 7 represent increasing levels of integrative complexity.

EXPERIMENTAL GROUPS

A total of 17 subjects scoring below 1.5 (mean of their two most complex responses) and referred to as the low-complexity or structurally concrete group, and 14 subjects scoring above 5 and referred to as the high-complexity or structurally abstract group were selected for the experiment.

In the experimental session, the subjects were given a list of 10 professions to rank according to intelligence required for membership within the profession. They were also asked to evaluate three colleges through use of a semantic differential consisting of three evaluative scales and two potency

and activity scales. Two of the colleges to be rated were authentic, and one was fictitious.

The subjects were then presented with information purported to represent the mean rankings the professions received from a group of 461 students attending the fictitious college. This communication was prepared individually for each subject and was based upon his own original rankings of the 10 professions. In the preparation of this counternorm message, the profession which the subject ranked highest and the three he ranked lowest were left unchanged. The six between these extremes were changed by a constant value of three ranks. The subjects were asked to rank on a 7-point scale the degree of divergence between their rankings and those of the fictitious college survey. Subjects could not refer to their original rankings, since these had been collected.

The subjects were then asked to rerank the three colleges through the use of the scales used previously in the experiment. In this way, approbation or depreciation of the fictitious college, as a result of the counternorm attributed to the college, could be ascertained. A divergence of at least 3 evaluative scale points between premeasure and postmeasure was necessary before a subject was said to approve or deprecate the source.

The subjects were then given a list of the professions and asked to reproduce the results of the fictitious survey. The difference between this reproduction and the original counternorm message provided a measure of assimilation or contrast. In order to determine which of these two opposing processes were utilized, it was necessary to refer to the subject's original rankings. When the subject changed the message in a direction away from his own rankings, this was regarded as an instance of contrast. If he altered the message so that it more closely agreed with his original rankings, then assimilation was considered to have taken place.

A list of the professions was then presented to the subjects who were asked to arrange them with respect to intelligence along an IQ scale provided on the page. The rankings resulting from this operation were then compared to the subject's original rankings. If the subject followed the influence of the counternorm message, it was assumed that the mechanism of acceptance had been utilized. If he changed his rankings in a direction opposite to that of the message, this shift was taken to be an indication of the process of rejection.

CONTROL GROUPS

It might be argued that the evidence of acceptance or rejection was merely an instance of a failure in recall on the part of some subjects. A group of 14 subjects was asked to rank the 10 professions and to estimate the average IQ of each of the groups. After 30 minutes, they were asked to arrange the professions along an IQ scale. There was perfect retention in all cases.

Further, a group of eight subjects was given a list purported to be the mean rankings the professions

TABLE 1

MANNER IN WHICH RESOLUTION PROCESSES
ARE UTILIZED IN REACTION TO A
COUNTERNORM MESSAGE

Groups	Ss using internally consistent processes	Ss using internally discordant processes
Concrete	16	1
Abstract	6	8

received from a group of 461 students of the fictitious college used in this study. After a 10-minute delay, the subjects were asked to reproduce these ratings. Again, there was perfect retention.

RESULTS

To determine whether the fictional college was accepted as real by the subjects, an analysis of variance was calculated, comparing the evaluative ratings which all three colleges had obtained in the pretest. No significant difference in evaluation was found. This leads us to conclude that the fictitious college was accepted by the subjects as a real institution.

The manner in which resolution processes were utilized by the two groups was then investigated. It was found that subjects operating at a low level of conceptual complexity tended to utilize processes in a manner defined earlier as internally consistent. The abstract group did not follow this trend. This divergence in method of utilization of conflict resolution processes between the structurally concrete and abstract groups was found, through a Fisher exact-probability test, to be significant at the $p < .005$ level (two-tailed).

An attempt was then made to determine whether one group utilized appreciably more dimensions than the other in reacting to attitudinal discrepancy. A Fisher exact-probability test was used in determining the statistical significance of these results. This analysis yielded a p value of less than .05 (two-tailed), illustrating the fact that the concrete group utilized resolution processes in a mutually exclusive manner significantly more often than the subjects in the structurally abstract group.

TABLE 2

NUMBER OF RESOLUTION PROCESSES UTILIZED IN
REACTION TO A COUNTERNORM MESSAGE

Groups	Ss utilizing 1 process	Ss utilizing more than 1
Concrete	11	6
Abstract	3	11

Finally, a point-biserial correlation was computed, investigating the relationship between conceptual complexity (the dichotomized variable) and the degree to which the subjects perceived the counternorm as diverging from their own opinions. Although this result failed to attain statistical significance ($r_{pt. bis} = -.19$, $N = 31$), the direction of this relationship is intriguing, especially in view of the fact that the actual divergence between the original opinion and the counternorm was exactly the same for all subjects.

DISCUSSION

Earlier in this study it was hypothesized that the individual operating at a low level of conceptual complexity would tend to utilize resolution processes in an internally consistent manner. Since inconsistency has been shown to be threatening to the structurally concrete individual (Harvey, Hunt, & Schroder, 1961; Schroder et al., 1966), there was good basis for this prediction. Consistent with the above investigations, the results of this study suggest that the typical method of coping with a counternorm or discrepant message for the integratively simple structure (i.e., attitude) is to utilize a single resolution process or group of internally consistent processes. This enables the conflict resulting from an attitudinally discrepant message to be most easily resolved, consequently reducing the threat of this material and resulting in a balanced cognitive state. This assumption is supported by the information presented in Table 1, which points to the fact that 16 of 17 subjects having structurally concrete attitudes utilized resolution processes in an internally consistent manner.

In Table 2, a comparison was made between subjects who utilized only one process in reaction to a counternorm message, and those using more than one. It was found that 11 of 14 subjects utilizing only one process were those classed as structurally concrete through Schroder and Streufert's (1963) Sentence Completion Test. Of the 17 subjects utilizing more than one process in response to the attitude discrepant message, 11 were classified structurally abstract. These results suggest that the work of Harvey et al. (1957) and Kelley and Woodruff (1956) needs to be extended. Both experiments suggested a negatively interdependent relationship between various mechanisms of conflict reduction, but accounted for this relationship on the basis of the "intersubstitutability" of these processes. As the results of the present study indicate, however, intersubstitutability is an acceptable ex-

planation only in the case of the structurally concrete individual, who tends to utilize processes in a mutually exclusive manner. When dealing with the more complex structure, negative interdependence results not from process intersubstitutability, but is rather a function of the multiple and internally discordant manner in which resolution processes are utilized. Consequently, the principle of intersubstitutability appears to hold only in the particular case in which simple attitude structures are involved. As the complexity of the attitude increases, a corresponding decrement in the adequacy of this rule is hypothesized.

In relating the results of this experiment to the work of Festinger (1957), the same general observations apply. The dissonance model implies that in the reduction of dissonance, an individual will utilize as many resolution processes as possible, and combine them in an internally consistent manner. On the basis of the results presented earlier, it is clear that only integratively simple subjects are bound by the condition of internal consistency. Over half of the individuals in the conceptually complex group utilized processes in an internally discordant manner—a fashion decidedly contrary to that expected on the basis of a dissonance-reduction principle.

Equally distressing from a dissonance standpoint is the fact that over 60% of the concrete subjects utilized only one process, when three were available. Although these results are certainly not consistent with a dissonance explanation, we do not mean to deprecate the dissonance principle, but rather to suggest that the scope and accuracy of dissonance research could be greatly increased if considerations of the conceptual complexity—and consequent differences in style of dissonance reduction—were taken into account.

REFERENCES

- ABELSON, R. P. Modes of resolution of belief dilemmas. *Journal of Conflict Resolution*, 1959, 3, 343-352.
- BIERI, J., & BLACKER, E. The generality of cognitive complexity in the perception of people and inkblots. *Journal of Abnormal and Social Psychology*, 1956, 53, 112-117.
- CARLSON, V. R. Individual differences in the recall of word association test words. *Journal of Personality*, 1954, 23, 77-87.
- COOPER, E., & JAHODA, M. The evasion of propaganda: How prejudiced people respond to anti-prejudice propaganda. *Journal of Psychology*, 1947, 23, 15-25.
- ERIKSEN, C. W. Defense against ego threat in memory and perception. *Journal of Abnormal and Social Psychology*, 1952, 47, 230-235.
- ERIKSEN, C. W. Psychological defenses and "ego-strength" in the recall of completed and uncompleted tasks. *Journal of Abnormal and Social Psychology*, 1954, 49, 45-50.
- FESTINGER, L. *A theory of cognitive dissonance*. Evanston, Ill.: Row, Peterson, 1957.
- HARVEY, O. J., HUNT, D. E., & SCHRODER, H. M. *Conceptual systems and personality organization*. New York: Wiley, 1961.
- HARVEY, O. J., KELLEY, H. H., & SHAPIRO, M. M. Reactions to unfavorable evaluations of the self made by other persons. *Journal of Personality*, 1957, 25, 395-411.
- HUNT, D. E., & SCHRODER, H. M. Assimilation, failure-avoidance, anxiety. *Journal of Consulting Psychology*, 1958, 22, 39-44.
- KELLEY, H. H., & WOODRUFF, C. Members' reactions to apparent group approval of a counternorm communication. *Journal of Abnormal and Social Psychology*, 1956, 52, 67-74.
- KELMAN, H. C. Attitude change as a function of response restriction. *Human Relations*, 1953, 6, 185-214.
- KELMAN, H. C. Compliance, identification, and internalization: Three processes of attitude change. *Journal of Conflict Resolution*, 1958, 2, 51-60.
- OSGOOD, C. E., & TANNENBAUM, P. H. The principle of congruity in attitude change. *Psychological Review*, 1955, 62, 42-55.
- SCHRODER, H. M., DRIVER, M. J., & STREUFERT, S. *Human information processing*. New York: Holt, Rinehart & Winston, 1966.
- SCHRODER, H. M., & STREUFERT, S. The measurement of four systems of personality structure varying in level of abstractness: Sentence completion method. Technical Report No. 11, Project NR 171-055, 1963, Princeton University, Contract NONR 1858 (12), Office of Naval Research.
- SHERIF, M., & HOVLAND, C. I. *Social judgment: Assimilation and contrast effects in communication and attitude change*. New Haven: Yale University Press, 1961.
- SUEDFELD, P. Attitude manipulation in restricted environments: I. Conceptual structure and response to propaganda. *Journal of Abnormal and Social Psychology*, 1964, 68, 242-247.

(Received March 7, 1966)

INCIDENTAL CONCEPT FORMATION AS A FUNCTION OF CREATIVITY AND INTELLIGENCE

PATRICK R. LAUGHLIN

Loyola University

College students were given an intentional concept-formation task as a cover for a later incidental concept-formation task. High, medium, and low creativity groups were then differentiated on the Remote Associates Test, and high, medium, and low intelligence groups were differentiated on the Terman Concept Mastery Test. Highly creative Ss were better on incidental concept formation than medium or low creative Ss, and medium better than low; while high, medium, and low intelligence groups did not differ. Differences between creativity groups remained significant with intelligence as a covariate, while previously nonsignificant differences between intelligence groups were significant with creativity as a covariate. It is suggested that the ability to form, retain, and utilize remote associations is the underlying process in both creativity and incidental learning, and that this ability is independent of high level verbal intelligence.

Mednick (1962) has presented an associative theory of the creative process. Essentially, the theory considers creativity the process of bringing previously unassociated ideas into contiguity so that previously unrealized relationships between them become apparent. Thus, the theory states that "any ability or tendency which serves to bring otherwise mutually remote ideas into contiguity will facilitate a creative solution [Mednick, 1962, p. 222]." Based on this associative theory of the creative process, Mednick has developed a Remote Associates Test (RAT) of creativity. This test consists of 30 sets of 3 words which are remotely associated or related to each other in some way which the subject must indicate by supplying a fourth word. A description of the test, with reliability and validity data, is given by Mednick (1962).

Incidental learning is fundamentally a situation in which apparently irrelevant ideas or relationships must later be associated or brought into contiguity. Thus, incidental learning involves the same underlying process postulated in Mednick's associative theory of creativity, and consequently a more creative individual should display better incidental learning than a less creative individual. Mendelsohn and Griswold (1964) suggested two possible factors underlying this hypothesized relationship between creativity and incidental learning: (a) The more creative individual may have greater perceptual sensitivity to the apparently irrelevant stimulation in an incidental learning situation; and/or (b) the more creative individual may retain the stimulation in a more accessible form for later problem solving in the incidental learning task. However, they did

not imply that creativity is merely better memory:

We do not mean to imply that the memory of high creatives is necessarily superior to that of other people, but rather, this notion suggests that what is retained, whether or not it appeared relevant to a given problem at the time of reception, is more available to such individuals during subsequent problem solving [Mendelsohn & Griswold, 1964, pp. 431-432].

In an experiment testing these two factors of perceptual sensitivity and accessibility for problem solving, Mendelsohn and Griswold had undergraduates memorize 25 words under the interference condition of playing another list of 25 words simultaneously on a tape recorder. Later they solved 30 anagrams, for which 10 of the solutions had been in the memorized list, 10 in the interference list, and 10 were new or neutral. High-creative subjects utilized both the cues from the memorized and interference lists to a greater extent than medium or low subjects, and medium to a greater extent than low. There were no differences in rote recall of the memorized or interference lists. These results were interpreted as "reflecting wider deployment of attention and less screening out of 'irrelevant' past experiences by high creatives during problem solving [Mendelsohn & Griswold, 1964, p. 431]."

Whatever the criterion measure, creativity has been shown to correlate positively with intelligence (Golann, 1963; McNemar, 1964). Thus, in research on creativity, intelligence differences must be controlled so that intelligence is not a more parsimonious explanation of differences attributed to creativity. Although Mendelsohn and

Griswold used scores on their 10 neutral anagrams as a covariate for the anagrams previously presented in the memorization and interference lists, their study did not include direct controls for intelligence. Thus, in the following study both creativity (Mednick, 1962) and intelligence (Terman, 1956) scores were obtained for the subjects. High, medium, and low groups were first differentiated on a basis of creativity, and then on a basis of intelligence. As a further control, intelligence was used as a covariate for groups differentiated on creativity, and creativity was used as a covariate for groups differentiated on intelligence. The following hypotheses were tested:

1. Performance on incidental concept formation will be positively related to degrees of creativity.
2. High, medium, and low intelligence groups will not differ on incidental concept formation. However, they will differ with analysis of covariance using creativity as covariate.

METHOD

Subjects

The subjects were 148 undergraduates in introductory and experimental psychology courses at Loyola University. They were run in groups of from 7 to 55.

Concept Materials

The materials for the intentional and incidental concept-formation tasks were taken from Underwood and Richardson (1956). These authors calibrated response dominance of word associations to each of 213 singly presented stimulus nouns by instructing 153 college students to give one word-association response to each noun. Responses were limited to sense impressions, for example, "hard," "long," or "red." The percentage of students giving a particular word association to a noun was considered an index of the associative strength of the response to the stimulus word. For example, to the stimulus noun BARREL, 72% of the students gave the association "round," 15% "woody," 6% "big," and 7% gave miscellaneous other responses.

Ten sets of six words per set were taken from the Underwood and Richardson materials. Four words in each set of six all evoked the same associative response in a high percentage of subjects, for example, the four words GLOBE, WHEEL, SPOOL, and BASEBALL all evoked the associative response "round." Thus the four words were considered exemplars of the common response or concept "round." The other two words in each set of six both evoked another common response unrelated to the concept exemplified by the first four words. For example, SAUERKRAUT and VINEGAR both evoked the same associative response "sour" and thus were exemplars of the concept "sour." The concept evoked by the four words was designated for the intentional concept-learning

task, and the concept evoked by two words was designated for the incidental concept-learning task. The six words were arranged randomly in each set. The 10 sets of six stimulus words, four words for the intentional concept-formation task and two words for the incidental concept-formation task, are presented in Table 1.

Procedure and Instructions

The procedure was designed to present both intentional and incidental concept-formation stimulus words together, with instructions to learn only the intentional concept. Each subject wrote his response in a booklet. The cover page of the booklet had the following directions, which the subject read while they were simultaneously read aloud:

Six words will be pronounced aloud. Four of these six words will go together in some way. These four words exemplify a concept. Listen carefully to the six words, and then figure out the concept or way in which four of the six words go together. Then write the concept word in the blank. For example, consider the six following words: "glue, paste, house, flypaper, rubber cement, gymnasium." The four words that go together in some way are "glue," "paste," "flypaper," and "rubber cement," because they are all "sticky." Thus, the concept is "sticky," and you should write "sticky" in the answer blank for the concept. Do not turn each page until you are instructed to do so.

There were four trials of the 10 sets of six words, each on a separate answer page. Within each trial both the order of the six words within a set and the order of the 10 sets were in a different random order. The stimulus words were read in a steady monotone with 10 seconds between sets of six words and sufficient time to turn the page between trials. After the final trial the directions on the last page of the booklet were:

Now, the four words that exemplified each concept are given below. For each of the four words try to recall the *other two* words that were *not* part of the concept. These two words, however, were also like each other in some way, and thus exemplified another concept. Write the two other words and the concept they exemplified below in the blanks provided.

The 10 sets of four words were presented in a new random order. Eight minutes were allowed for recall of the two incidental words and for incidental concept formation. Two to 4 weeks later subjects were administered the RAT (Mednick, 1962) and the Terman (1956) Concept Mastery Test.

RESULTS

In order to get as close as possible a breakdown of subjects into high and low quartiles, 32 subjects who scored 22 or above (maximum =

TABLE 1
STIMULUS WORDS FOR INTENTIONAL AND INCIDENTAL CONCEPT FORMATION

Set	Intentional concept	Incidental concept	Set	Intentional concept	Incidental concept
1	Round (95) GLOBE (95) WHEEL (91) SPOOL (74) BASEBALL (70)	(Soft) VINEGAR (68) SALERKAUT (41)	6	(Dark) NIGHT (90) DUNGEON (67) CLOSET (64) TUNNEL (54)	(Slimy) FEL (68) LIZARD (51)
2	Salty AMMONIA (88) MANIST (83) GARBAGE (80) SKUNK (78)	(Cold) FROST (54) ICICLE (45)	7	(Large) AULICORUM (84) ELEPHANT (83) MANSION (83) OCEAN (33)	(Sweet) SUGAR (82) HONEY (49)
3	Sharp KNIFE (84) HATCHET (77) FANG (75) TACK (64)	(Black) COAL (85) TELEPHONE (65)	8	(Shiny) JEWEL (67) DIAMOND (65) ALUMINUM (59) BADGE (32)	(Brown) TOBACCO (83) CHOCOLATE (61)
4	(White) MILK (83) TEETH (72) SNOW (71) IVORY (65)	(Soft) PILLOW (87) VELVET (67)	9	(Red) BLUSH (96) BIFT (87) CHERRY (77) APPLE (67)	(Hard) STONE (63) KNUCKLE (62)
5	(Green) SPINACH (90) GRASS (88) IVY (88) LAWN (77)	(Spicy) GINGER (40) CLOVE (32)	10	(Small) ATOM (87) FLEA (86) GERM (84) GNAT (76)	(Yellow) CANARY (82) DANDELION (85)

Note.—Numbers in parentheses indicate percentage of subjects giving response as first association to stimulus word in Underwood and Richardson (1956).

30) on the RAT comprised the high creativity group, 34 who scored 12 or below, the low creativity group, and the remaining 82, the medium creativity group. Means for the high, medium, and low creativity groups were 24.03, 17.29, and 9.24, respectively. Numerically comparable high, medium, and low intelligence groups were comprised of the highest 32 subjects on the Terman Concept Mastery Test, the lowest 34, and the remaining 82. Means for the high, medium, and low intelligence groups were 110.78 (maximum = 180 with a right minus wrong correction procedure), 66.98, and 30.26. Creativity and intelligence correlated .48 over all subjects; .18, .16, and .24 within high, medium, and low creativity groups; .19, .20, and .50 within high, medium, and low intelligence groups.

Scores on incidental concept formation and intentional concept formation were analyzed in four ways: (a) analysis of variance for creativity groups, (b) analysis of covariance for creativity groups with intelligence as covariate, (c) analysis of variance for intelligence groups, (d) analysis of covariance for intelligence groups with creativity as covariate. Means for incidental concept formation and intentional concept formation for groups differentiated on creativity and for groups differentiated on intelligence are

given in Table 2. Summary tables for analyses of variance and covariance for incidental and intentional concept formation for groups differentiated on creativity and on intelligence are given in Table 3.

TABLE 2
MEAN INCIDENTAL CONCEPT FORMATION AND INTENTIONAL CONCEPT FORMATION FOR GROUPS DIFFERENTIATED ON CREATIVITY AND FOR GROUPS DIFFERENTIATED ON INTELLIGENCE

Creativity	Incidental concept formation		Intentional concept formation	
	M	Adjusted M (Intelligence as covariate)	M	Adjusted M (Intelligence as covariate)
High (n = 32)	4.09	3.69	37.72	37.39
Medium (n = 82)	3.03	3.03	35.76	35.77
Low (n = 34)	1.94	1.84	35.15	36.11
Intelligence	(Creativity as covariate)		(Creativity as covariate)	
	M	Adjusted M	M	Adjusted M
High (n = 32)	3.13	1.99	37.47	37.43
Medium (n = 82)	3.27	3.36	35.48	35.49
Low (n = 34)	2.32	2.52	36.06	36.57

Note.—Maximum incidental concept formation is 10.00; maximum intentional concept formation is 40.00.

TABLE 3

ANALYSES OF VARIANCE AND COVARIANCE FOR INCIDENTAL AND INTENTIONAL CONCEPT FORMATION FOR CREATIVITY AND INTELLIGENCE GROUPS DIFFERENTIATED ON CREATIVITY AND FOR GROUPS DIFFERENTIATED ON INTELLIGENCE

	Incidental concept formation						Intentional concept formation					
	Analysis of variance			Analysis of covariance			Analysis of variance			Analysis of covariance		
	MS	df	F	MS	df	F	MS	df	F	MS	df	F
Creativity	36.14	2	10.12***	27.12	2	7.45***	120.39	2	3.85*	37.64	2	2.88
Error	3.57	145		3.64	142		2281.80	145		15.83	142	
Intelligence	10.95	2	2.80	11.23	2	3.19*	45.74	2	2.90	18.47	2	1.19
Error	3.91	145		3.51	142		15.78	145		15.52	142	

* $p < .05$.

*** $p < .001$.

Incidental Concept Formation

The mean numbers of correct incidental concepts (maximum = 10) for the high, medium, and low creativity groups were 4.09, 3.03, and 1.94. Overall analysis of variance resulted in a significant difference at the .001 level, $F(2, 145) = 10.12$. The differences remained significant at the .001 level with analysis of covariance using intelligence as covariate, $F(2, 142) = 7.45$. Adjusted means for high, medium, and low creativity groups were 3.69, 3.03, and 1.84. On adjusted means high creatives differed from low creatives at the .001 level, $t(64) = 4.53$, and medium creatives differed from low creatives at the .005 level, $t(114) = 3.62$. High and medium creatives did not differ significantly.

The mean number of correct incidental concepts for the high, medium, and low intelligence groups were 3.13, 3.27, and 2.32. Overall analysis of variance was nonsignificant, $F(2, 145) = 2.80$. With analysis of covariance using creativity as covariate the difference was significant at the .05 level, $F(2, 142) = 3.19$. On adjusted means medium intelligence groups performed better than high intelligence at the .001 level, $t(112) = 3.38$, and medium intelligence better than low at the .01 level, $t(114) = 2.56$. High and low intelligence groups did not differ significantly.

Thus, groups differentiated on a basis of creativity differed on incidental concept formation, while groups differentiated on a basis of intelligence did not. Further, the difference between creativity groups remained significant with intelligence as a covariate, while previously nonsignificant differences between intelligence groups were significant with creativity as a covariate.

Intentional Concept Formation

The mean number of correct intentional concepts formed for the high, medium, and low

creativity groups was not significant when intelligence was controlled.

In addition to total score over the four intentional concept-learning trials, the final trial alone was also analyzed as a measure of final learning. Means for high, medium, and low creativity groups were 9.61, 9.41, and 9.29 (maximum = 10), while means for high, medium, and low intelligence groups were 9.69, 9.27, and 9.47, respectively. Neither the differences between the three creativity groups nor the intelligence groups were significant, $F(2, 145) = 1.37$, $F(2, 145) = 2.86$.

Thus, groups differentiated on a basis of creativity differed on intentional concept learning over four trials but not on the final trial, while groups differentiated on a basis of intelligence did not differ either over four trials or on the final trial. However, the difference between creativity groups over four trials did not remain with intelligence as a covariate.

Correlations between Measures

Over all subjects, creativity scores on the RAT correlated .35 with incidental concept formation and .25 with intentional concept formation. Intelligence scores on the Concept Mastery Test correlated .20 with incidental concept formation and .16 with intentional concept formation. Finally, the recall of incidental stimuli correlated .91 with incidental concept formation, and consequently the recall data were not analyzed apart from incidental concept formation.

DISCUSSION

The deduction from Mednick's (1962) associative contiguity theory of creativity was supported, as high creatives showed better incidental concept formation than medium and low creatives, and medium creatives better incidental concept formation than low creatives.

These results seem to indicate creativity and incidental learning involve the same basic underlying process of bringing apparently irrelevant or unrelated ideas into contiguity or association so that the relationship between them is strengthened. In the present experiment the individual is under the intentional learning set to find the relationship or concept between four of six words, which requires that he form a common association between the four words. However, he simultaneously is presented with two other incidental words, unrelated to the four intentional words but related to each other. Thus, he may likewise form an association between these two words. The more creative person apparently is more proficient at forming and retaining this incidental association, or remote association in Mednick's (1962) terms, and in utilizing it in later incidental concept formation. Thus, the ability to form, retain, and utilize remote associations seems to be the underlying process in both creativity and incidental learning. This has been demonstrated with two different types of prob-

lem-solving tasks, anagrams by Mendelsohn and Griswold (1964) and concept formation in the present study. Furthermore, at least in a college population, this ability was demonstrated to be relatively independent of differences in high level verbal intelligence in the present study.

REFERENCES

- GOLANW, S. E. Psychological study of creativity. *Psychological Bulletin*, 1963, **60**, 548-565.
 McNEMAR, Q. Lost: Our intelligence? Why? *American Psychologist*, 1964, **19**, 871-882.
 MEDNICK, S. A. The associative basis of the creative process. *Psychological Review*, 1962, **69**, 220-232.
 MENDELSON, G. A., & GRISWOLD, B. B. Differential use of incidental stimuli in problem solving as a function of creativity. *Journal of Abnormal and Social Psychology*, 1964, **68**, 431-436.
 TERMAN, L. M. *Manual for Concept Mastery Test*. New York: Psychological Corporation, 1956.
 UNDERWOOD, B. J., & RICHARDSON, J. Some verbal materials for the study of concept formation. *Psychological Bulletin*, 1956, **53**, 84-95.

(Received January 31, 1966)

Journal of Personality and Social Psychology
 1967, Vol. 5, No. 1, 119-122

EXPERIMENTAL MODIFICATION OF DRAMATIC PLAY¹

HELEN R. MARSHALL AND SHWU CHING HAHN

New Mexico State University

This experiment attempted to determine a cause of correlations between parent and child behaviors obtained earlier by Marshall. 12 matched triads of preschool children were trained in doll-play fantasy, or in use of toys, or given no training. Frequency of dramatic play before and during training was the dependent variable. The results showed that if an adult engages in fantasy play with a child, enacting topics commonly used in children's dramatic play with peers, the child will increase the frequency of his dramatic play with peers. Training in dramatic play was the critical factor, not the warmth and attention confounded with training in such play.

In an earlier investigation (Marshall, 1961), frequency of dramatic (imaginative) play with peers correlated positively with acceptance by peers, with number of friendly interactions, and with independence of teachers during play. These often are described as desired characteristics for preschool children. The demonstration of an antecedent of dramatic play would suggest a way to encourage these desired behaviors.

By experimentally increasing the amount of children's dramatic play with peers, this experi-

ment attempted to test a cause of correlations between parent behavior and dramatic play, also obtained in Marshall's investigation. The obtained correlations were between the number of dramatic play topics the child had talked about with adults, and the frequency of the child's dramatic play at nursery school. For example, in the sample of 49 girls, frequency of dramatic play suggestions correlated .44 with fathers' reports, .43 with mothers' reports, and .52 with parents' joint reports of topics talked about with other adults. The correlations suggested a testable hypothesis: When an adult talks with preschool children about the topics commonly used in children's dramatic play with peers, the chil-

¹ This research was supported, in part, by a grant from the Research Center, New Mexico State University.

dren will increase the frequency of their dramatic play with peers at nursery school.

If that adult is an experimenter, however, he cannot expect his words to influence children as much as those of the parent. A more dramatic situation, such as doll play, seems essential if verbal training is to exert an effect in a few short sessions. Marshall and Doshi (1965) found a close relation between the frequency of fantasy in doll play and that in dramatic play with peers. Thus, training in doll-play fantasy could be expected to transfer to dramatic play with peers. These ideas modify the specific hypothesis to be tested: If an adult engages in fantasy play with a child, enacting topics commonly used in children's dramatic play with peers, the child will increase his dramatic play with peers.

The hypothesis was tested by comparing the frequency of dramatic play with peers before and during several sessions of doll-play fantasy.

If an adult trains a child, by playing with him, the training is confounded with warmth and attention. However, the warmth and attention given in doll-play fantasy can be given equally well during the assembly of plastic blocks or the fitting of puzzle parts. This factor, the interaction with an adult in play, was studied by employing a matched control group, which was given sessions of play with eye-hand coordination and science-concept toys. The sessions were of the same duration, and the experimenter talked as much and as warmly as in the doll-play sessions. A second matched control group was given no training.

METHOD

Sample

Twelve triads of preschool children were matched in sex, age, ordinal position in the family, status of father's occupation, and for length of attendance at nursery school by the director of their school. Children within a triad were assigned at random to experimental groups. Nine children attended the nursery school of Drexel Institute of Technology, Philadelphia. The other 27 children attended the New Mexico State University Nursery School-Kindergarten.

There were seven triads of girls and five triads of boys. Age ranged from 33 to 66 months at the initiation of training, and the mean age was 48.7 months. Fathers' occupations were professional, business managerial, and student.

Training

Each child in the doll-play group was given at least four training sessions of 15 minutes duration. His sessions were spaced a day apart, except when the child was absent. Then, an additional session preceded the recording of nursery school behavior. The experimenter (the first author)² has had many

years of experience with preschool children, and usually has been described as "warm" in all his interactions with them. The experimenter spoke to each child at least once a minute during training. Each child in the toy-training group was given similar sessions.

At the first doll-play session, the experimenter and child played with a Flagg doll family, their house and station wagon, and a gas station. The experimenter initiated most activities for the family. New dolls included grandparents, other families, a mailman, a policeman, a doctor, and a nurse. New places included a zoo, a farm, a doctor's office, and a vegetable stand. Only New Mexico children had a stable, while only Philadelphia children had a restaurant.²

At the first toy-training session, the experimenter and child played with as many as five toys (differing with the child's age) as the child's attention span permitted. Two new toys were introduced at each succeeding session. All toys were different from those available to the children at nursery school.³ The experimenter's verbalization focused on number, color, form, etc., and avoided fantasy statements.

In Philadelphia, training sessions were held in a room in the school not used for play. In New Mexico, a room in a building a block from the school was furnished for these sessions. Philadelphia children selected a reward toy from an array of dime toys after each session. The trip to another building seemed as effective a reward for New Mexico children.

Behavior Records

Time-sample records of behavior were taken during free-play periods in a rotating order for the three children. No more than eight records per child were taken on any day, and the mean number per day was 4.2.

At least 40 2-minute records had been collected on each child before training was initiated. The most recent 40 records for each child furnished the "before" training data. Records then were taken during training until there were 40 for each child, and these furnished the "during" training data. Before and during training records were taken within the same 6 calendar weeks for a particular triad of children.

² All doll-play equipment was obtained from Creative Playthings and Community Playthings, except the station wagons and vegetable stands, which were purchased locally.

³ The following toys, listed by their commercial names, were used at both schools: Giant Magnifier, Fruits that Grow on a Tree, Vegetables that Grow in the Ground, Toymaker, Flexible Mirror, and Birds in a Tree puzzle. Additional toys in New Mexico were: Keys of Learning, Stack'em Chairs, Fit-a-Space, Petal Craft, Feathered Friends that Live Together, and a basket of plaster animals with moving heads and tails. Additional toys in Philadelphia were: two bicycle horns, a Chinese chime, Rig-a-Jig, Kitten in a Keg, Ten Second Timer, and Comic Peg Board.

Most of the records were taken by the second author, after establishing 85% agreement in entries for 10 consecutive records with the senior author. Two graduate students took two-thirds of the records on six children, also after establishing 85% agreement with the senior author. The senior author took 10-20 records on each child. Only the senior author knew the assignment to training groups and the behaviors pertinent to the hypothesis.

The frequency of dramatic play with peers was measured by two scores, whose relations to other variables are described in Marshall (1961). One, dramatic play language, is a count of the number of records in which the child used friendly or neutral language as if he were the person, animal, or thing that he was imitating in dramatic play. The other, dramatic play aggression, is a count of the number of records in which the child used aggressive words or actions to carry out his role in dramatic play. Aggressive responses include attack, threat, interference with activity, snubbing, and name calling.

Four other types of interaction with peers were recorded and analyzed as a check on the adequacy of matching within triads, and on possible generalization of the training. The types of interaction were associative play, friendly approach or response, conversation, and aggression, as developed by Marshall and McCandless (1957). Scores are the number of children in 40 records with whom the child had that type of interaction.

RESULTS

Before training, the subjects were well matched within triads in regard to age and the six behavior scores. In 3×12 Treatment \times Triads analyses of variance (appropriate because of the matching within triads), only the F for friendly approach interactions was significant (.01 level). By Duncan's new multiple-range test (Edwards, 1964), the toy-training control group had more (.01 level) friendly approaches to other children before training ($M = 53.4$) than the doll-play group (45.8), but did not differ from the no-training group (49.1). Friendly approach includes talk as self, as well as dramatic play language. Because significant differences were not obtained for dramatic play language, this difference probably is due to more talk as self.

During training, the doll-play group increased the frequency of their dramatic play with peers, as is shown in Table 1, and this increase did not occur in the two control groups. The F for dramatic play language was 11.62, significant at the .001 level. By Duncan's test, the doll-play mean differed at the .001 level of significance from the no-training mean, and at the .005 level from the toy-training mean. The difference between means for the two control groups was not significant.

The F for dramatic play aggression was 4.96, significant at the .025 level. By Duncan's test,

TABLE 1

MEAN DIFFERENCES (DURING TRAINING-BEFORE TRAINING) IN DRAMATIC PLAY SCORES OF TRAINING GROUPS

Scores	Doll-play training	Toy training	No training
Dramatic play language	+7.1	-0.1	-1.5
Dramatic play aggression	+3.7	+0.7	+0.4

the doll-play mean differed from the no-training mean at the .01 level of significance, and from the toy-training mean at the .05 level. Again, control-group means did not differ significantly.

These data support the hypothesis: If an adult engages in fantasy play with a child, enacting topics commonly used in children's dramatic play with peers, the child will increase his dramatic play with peers. They also indicate that training in dramatic play is the critical factor, not the warmth and attention confounded with training in such play.

All doll-play children had more frequent dramatic play language during training, as is shown in Table 2, and all but one had more frequent dramatic play aggression. Children in the two control groups lacked consistency in the direction of difference. The size of the difference in the doll-play group ranged from +3 to +17 for dramatic play language, and from -1 to +9 for dramatic play aggression.

During training, there was little generalization to other social interactions with peers. The only significant F was for differences in associative play. The doll-play group had 10.5 more associative play interactions with peers. By Duncan's test, this differed at the .005 level from the toy-training mean (-1.75), and at the .001 level from the no-training mean (-3.5). Dramatic play with peers is a component of associative play, defined by Marshall and McCandless (1957) as sharing a common interest, activity, or goal.

TABLE 2

NUMBER IN TRAINING GROUPS WITH EACH DIRECTION OF DIFFERENCE IN DRAMATIC PLAY SCORES

Training group	Increase	Decrease	Same
Dramatic play language			
Doll-play training	12	0	0
Toy training	5	7	0
No training	3	8	1
Dramatic play aggression			
Doll-play training	11	1	0
Toy training	6	3	3
No training	6	2	4

DISCUSSION

The results of this experiment support the hypothesis that adult-child conversation about topics commonly used in dramatic play is an antecedent of more frequent dramatic play with peers. They suggest a basis for the correlations obtained earlier by Marshall (1961).

In this experiment, adult initiation of activities in doll play fostered children's self-directed fantasies with peers. The data suggest it was exposure to ideas that fostered use of ideas in play, and not adult warmth and attention, or lack of exposure to adult conversation. Three aspects of the exposure to ideas may have contributed to the increase in dramatic play. First, verbal suggestions may have placed more value on the ideas for the child. Second, acting-out ideas furnished a model for the child to imitate. Children imitated aggressive and postural movements of adult models in experiments of Bandura and co-workers (e.g., Bandura, Ross, & Ross, 1963). Finally, there were two types of reinforcement for the child's contributions. The experimenter accepted and furthered the child's ideas, and the child enjoyed playing with an adult.

Adult warmth and attention were not varied in this experiment, but merely controlled. Hence, the possibility that this factor might interact with training has not been ruled out. This could be studied by having the experimenter interact coldly during doll play with some children, and warmly with others.

The results suggest a way to encourage desired behavior in peer groups. If parents and adults talk with the child about topics the child can use in dramatic play, the child may be expected

to increase his dramatic play with peers. In turn, this should benefit the child's adjustment in his peer group. Correlations obtained in Marshall's earlier investigation indicate that more dramatic play is accompanied by better acceptance by peers, more friendly interactions with peers, and greater independence of teachers during play.

Further experimentation varying the number of exposure to fantasy training might disclose minimum and maximum effects of training. Study might be made also of the effect of stimulus characteristics, such as realism versus abstractness in dolls and houses, or of sex appropriateness of the materials. Child characteristics that might interact with training include similarity of conversational styles of experimenter and parent, frequency of dramatic play at initiation of training, and the child's memory skills.

REFERENCES

- BANDURA, A., ROSS, D., & ROSS, S. A. Vicarious reinforcement and imitative learning. *Journal of Abnormal and Social Psychology*, 1963, 67, 601-607.
- EDWARDS, A. L. *Experimental design in psychological research*. (Rev. ed.) New York: Holt, Rinehart & Winston, 1964.
- MARSHALL, H. R. Relations between home experiences and children's use of language in play interactions with peers. *Psychological Monographs*, 1961, 75(5, Whole No. 509).
- MARSHALL, H. R., & DOSHT, R. Aspects of experience revealed through doll play of preschool children. *Journal of Psychology*, 1965, 61, 47-57.
- MARSHALL, H. R., & MCCANDLESS, B. R. A study in prediction of social behavior of preschool children. *Child Development*, 1957, 28, 149-159.

(Received February 4, 1966)

Journal of Personality and Social Psychology
1967, Vol. 5, No. 1, 122-126

EFFECTS OF A RATIONAL APPEAL AND OF ANXIETY ON CONFORMITY BEHAVIOR

KAY H. SMITH AND BARRIE RICHARDS

Brigham Young University

This study involved a test of the hypothesis that conformity under group pressure is a form of ego defense against anxiety. 80 Ss selected from 2 introductory psychology classes who differed in pretest levels of anxiety (Taylor Manifest Anxiety scale—MA scale) and tendency to conform (Over-all Agreement Score—OAS) were subjected to group pressure. Ss from 1 class were given an anticonformity lecture shortly before the group-pressure session. Conformity was positively related to MA scale. Those who heard the lecture conformed more than those who had not. The effects of the lecture interacted with OAS. Low OAS Ss who had heard the lecture showed high conformity, while similar Ss who had not heard it showed low conformity. The results, in general, supported the ego-defense explanation of conformity.

The phenomenon of conformity, as a result of group pressure has become an established topic of study in social psychology. Many studies have been conducted attempting to discover the facets

personality which are associated with greater or lesser tendencies to conform. Still others have studied variations in factors in the situation or in the group aimed at increasing or decreasing group pressure. Studies have been done manipulating the subject's resistance to group pressure by varying his evaluation of the other group members relative to himself (Samelson, 1957; Smith, 1961), but little has been done in attempting to strengthen the subject's willingness or ability to resist group pressure in general. Blake and Mouton (1961) concluded, as a result of a survey of the literature on conformity:

Though it is yet to be demonstrated experimentally, it is also probable that an individual would be aided in maintaining independence through acquaintance with an understanding of the psychological forces producing conformity [p. 23].

Breger and Ruiz (1963) conducted a study to investigate whether an understanding of the psychodynamics of conformity would indeed produce greater resistance to conformity pressure. They also investigated the effects of a rational appeal to resist conformity pressures. The basic postulate under which the study was conducted was that conformity to group pressure is in large part a defense against the anxiety generated by being at odds with the group. This kind of ego defense against anxiety would operate largely on the unconscious level and therefore not be subject to rational appeal nor to increased knowledge or understanding. The anxiety is a result of the person's feeling that nonconformity is a form of aggression against the group and might result in his being cut off from the group. When the person conforms, he receives an immediate reduction in anxiety which, according to reinforcement theory, would be rewarding and tend to cause this behavior to be repeated in like situations. There is evidence that anxiety is positively related to susceptibility to group pressure (cf. Hoffman, 1957; Lawson & Stagner, 1957; Sherif & Harvey, 1952), although some contradictory results have also been obtained (Holder, 1958; Janis, 1954; Mangan, Quartermain, & Vaughan, 1960).

As predicted by Breger and Ruiz and contrary to the speculation of Blake and Mouton, the group who had received information about group-pressure research findings and procedures shortly before being subjected to experimentally induced group pressure did not behave in a significantly different way from the control group. A surprising finding, however, was that the group which had been subjected to an anticonformity lecture prior to their group-pressure session conformed more frequently than the members of either of the other two groups.

Breger and Ruiz offered the post hoc suggestion that individuals with tendencies to conform become anxious when they hear arguments advocating nonconformity. This anxiety is further heightened when they are faced with group pressure to conform. The ego-defense hypothesis of conformity would predict that greater anxiety would result in a greater tendency to conform.

The study reported here had two major purposes: (a) to replicate and further substantiate the results of the Breger and Ruiz study, and (b) to obtain data in support of the anxiety-defense explanation of these findings. The design of the study was essentially that of Breger and Ruiz with several modifications. The more important of these modifications was the addition of measures of conformity and anxiety. In addition to this, the condition involving explanation of conformity studies was not repeated.

The following specific hypotheses were investigated in the study:

Hypothesis 1. There is a positive relationship between anxiety and the degree of conformity in response to group pressure.

Hypothesis 2. There is a positive relationship between agreeing-response tendency and the degree of conformity in response to group pressure.

Hypothesis 3. Persons who have heard an anti-conformity appeal conform to a greater degree than persons who have not heard such an appeal.

Hypothesis 4. Persons with high anxiety conform more often in a group-pressure situation in response to items relating to information than to items relating to attitudes toward aggression, while this is not true for persons with low anxiety.

Hypothesis 5. There is a negative relationship between awareness of conformity pressure and degree of conformity in response to group pressure.

METHOD

Variables

Anxiety. The scale which was used to measure anxiety was the Taylor Manifest Anxiety scale (MA scale—Taylor, 1953). The short 50-item form was used (Bendig & McCreary, 1954). The subjects selected for the low and high anxiety groups obtained MA scale scores from 0 to 10, and from 20 to 50, respectively.

Tendency to conform. An ideal research design would require both a pretreatment and a posttreatment measure of conformity in response to group pressure. Since this was not practically possible, the Over-all Agreement Score (OAS) developed by Couch and Keniston (1960) was used for the pretreatment conformity measure. Although this scale is new, the descriptions given by Couch and Keniston of the yeasayer and the naysayer seemed to parallel closely those of yielders and resisters in group-

pressure studies (Asch, 1952). The low and high tendency to conform subjects obtained OAS scores from 0 to 3.29, and from 3.45 to 4.30, respectively.

Anticonformity lecture. The subjects in the experimental group heard a lecture by a senior advocating strongly a return to personal values, resistance to pressures toward conformity from the community, and individual thinking in forming opinions and making value judgments. The student lecturer was serving as the chairman of the student Academic Emphasis Committee. This committee had taken conformity as their theme for the year and had brought to the campus such men as David Reisman, Eric Hoffer, and William C. Barrett to speak on the subject. The student was very sincere, and it was evident from the reaction that the class members were considerably aroused by his appeal. He was introduced by the instructor as an invited guest speaking on a topic relevant to the course content. No connection was made between him and the study in which the students were participating.

Pretreatment questionnaire. The pretreatment questionnaire consisted of 60 true-false items. Of these items, 30 were statements of fact and 30 were statements of attitude toward the expression of aggression. Ten items were chosen in each category to serve as the conformity stimuli in the group-pressure situation. The information items chosen had at least 70% agreement and the aggression items, 75% agreement by the subject pool. Other items for which agreement was near 50% were used as filler items.

Group-pressure situation. In the group-pressure situation, five subjects were seated in booths similar to those used by Crutchfield (1955). In each booth was a panel of five lights. In front of the light corresponding to the position of the booth was a pushbutton-type switch. In checking out the equipment prior to the actual testing, the subject became aware that when he pressed his switch, the corresponding light went on in his booth, and he assumed that the lights in the same position in the other booths also were lit. This assumption seemed logical since lights went on in his booth when subjects in other booths were asked to press their switches. Actually, all the lights were under the control of the experimenter except the one corresponding to the switch and a similar light on the experimenter's panel which were under the control of the subjects. The subjects were instructed to listen to the items as they were played from a pre-recorded tape and to press their switches in the interval between items if they felt the item was true or to refrain from pressing if they felt that it was false. The experimenter then created varying impressions as to the responses of the other subjects by lighting the other lights in each booth according to a prearranged pattern. For the conformity items—which were all worded so that the majority response was false—all four of the other lights were lit, indicating responses opposite to the majority response on the pretreatment questionnaire, while for the filler items from zero to three lights were lit in

a random fashion. This procedure made unnecessary the subterfuge usually required to make the subject think he is always the last to respond, since in virtually all cases his pretest response was false, and thus he was extremely unlikely to press his switch for a true response until he had seen the other lights. On none of the conformity items did the subjects respond before they had seen the other lights.

Posttreatment questionnaire. After the group-pressure session, each subject filled out a short questionnaire, aimed at determining whether or not the subject had suspected the deception employed, the degree to which he was aware of a discrepancy between the way he had answered the items in the pretreatment questionnaire and the way others were answering the items in the group situation, and his general interest in and enjoyment of the experiment.

Sample

The subjects were students in two large introductory psychology classes, both taught by the same instructor. The 379 students in these two classes formed the subject pool. From this pool 10 subjects in each class were chosen in each of the following categories: low *MA* scale score, low OAS; low *MA* scale score, high OAS; high *MA* scale score, low OAS; and high *MA* scale score, high OAS. This yielded a total of 80 subjects. The subjects in one class were given the anticonformity lecture while the others served as controls. The final result was a three-dimensional design with 8 cells and 10 subjects in each cell.

Procedure

The experiment was carried out in three phases. In the first phase the *MA* scale, OAS, and pretreatment questionnaire were administered to all members of both classes. These scales were scored, the subjects selected, and the conformity scale made up. After a period of about 3 weeks, the anticonformity lecture was given to the experimental group. This comprised the second phase of the experiment. Prior to the delivery of the lecture to the experimental group, the subjects of the control group participated in Phase 3, the group-pressure session. The post-treatment questionnaire was administered as part of Phase 3 immediately following the group-pressure session. The group-pressure sessions for the experimental subjects were carried out either on the day the lecture was given or on the following morning.

RESULTS

Conformity scores were computed simply by counting the number of conformity items on which the subject changed the answer he had given on the pretest to conform with the supposed consensus of the other members of his five-person group. The summary table giving the results of the analysis of the conformity scores is shown in Table 1.

The results shown in Table 1 indicate support

for Hypotheses 1 and 3. Hypothesis 2 was not supported. Although no interaction was predicted, OAS interacted significantly with the nonconformity lecture. This result was of particular interest, since the main effect of the OAS variable was not significant. The source of this interaction was sought by collapsing cells along the *MA* scale dimension and then making all possible comparisons between the remaining cells by means of Duncan's multiple-range test. The only significant difference between means resulting from this comparison was that between the mean of the low OAS nonlecture group (1.74) and the low OAS lecture group (3.64). This difference was significant at the .05 level. Both the high OAS cells had means of 2.92. This result would seem to indicate that the lecture had the effect of increasing conformity for the naysayers, but not for the yeasayers. Why this would happen is difficult to say. Possibly the yeasayers do not have a great deal of anxiety about this habitual behavior and thus were not particularly upset by the anticonformity lecture. If this is a reliable result, it certainly would obscure the main effect of the relationship between OAS scores and conformity under group pressure, and thus account for the nonsignificant *F* ratio.

Further light is shed upon this subject by inspection of the mean conformity scores of the eight cells of the original three-way analysis. These are shown in Table 2, in descending order of magnitude. All possible mean differences were analyzed by Duncan's multiple-range test. Of the 28 mean pairs thus tested, only 4 showed significant differences. These were the differences between the mean of the high *MA* scale score, low OAS lecture group and the four lowest means in the table. This shows that the interaction is a result primarily of the scores in this

TABLE 2

CELL MEANS FROM THE THREE-WAY ANALYSIS

Cells	<i>M</i>
High <i>MA</i> scale score, low OAS, lecture	4.78
High <i>MA</i> scale score, high OAS, nonlecture	3.56
High <i>MA</i> scale score, high OAS, lecture	3.10
Low <i>MA</i> scale score, high OAS, lecture	2.90
Low <i>MA</i> scale score, low OAS, lecture	2.50
Low <i>MA</i> scale score, high OAS, nonlecture	2.28
High <i>MA</i> scale score, low OAS, nonlecture	1.78
Low <i>MA</i> scale score, low OAS, nonlecture	1.70

high *MA* scale score, low OAS lecture group, since the mean of the low *MA* scale score, low OAS lecture group was not significantly higher than that of either of the low OAS nonlecture groups.

Hypothesis 4 states that persons with high anxiety will conform more on information items than on aggression items, while this will not be true for low-anxiety persons. This prediction is based on the reasoning that nonconformity is perceived by the highly anxious person as unacceptable aggression against the group and that conformity avoids this aggression, thus reducing anxiety. To conform to items advocating free expression of aggression, then, would be to reinstate the problem. Analysis revealed that the mean differences between conformity to the two types of items produced *t*'s significant beyond the .05 level for both the high *MA* scale score and low *MA* scale score groups. While it is possible that this might give some support to the general hypothesis that conformity is a reaction against aggression for both high- and low-anxiety subjects, the interpretation is complicated by significantly greater pretest agreement for aggression items. This may mean that, on the average, the subjects were less sure of their answers on the information items and thus more susceptible to conformity pressures.

The fifth hypothesis stated that there would be a negative relationship between awareness of conformity pressure and degree of conformity in response to group pressure. The subjects had been asked a question regarding their awareness that the responses of the rest of the group differed from their own pretest answers to the items. Dichotomizing the conformity scores at the median and dividing responses to this question into those showing considerable awareness and those showing some or little awareness produced a $2 \times 2 \chi^2$ of 2.57 which is not significant at the .05 level. It was felt that possibly an analysis of only the high *MA* scale score lecture and low *MA* scale score nonlecture groups would more

TABLE 1

ANALYSIS OF VARIANCE OF CONFORMITY SCORES

Source	<i>df</i>	<i>MS</i>	<i>F</i>
<i>MA</i> scale (A)	1	18.29	4.51*
Lecture (B)	1	19.57	4.83*
OAS (C)	1	1.48	.36
A \times B	1	1.60	.39
A \times C	1	.78	.19
B \times C	1	16.59	4.09*
A \times B \times C	1	6.51	1.60
Within cells	72*	4.05	—

* Due to missed appointments, only 72 of the original 80 subjects actually participated in the experiment. As a result, data were generated at the cell mean for 8 subjects in five cells in order to achieve equality of cell *N*s for the analysis of variance. The mean values reported in Table 2 and in the body of the article are based on the actual *N*s.

* $p < .05$.

clearly reveal the possible impact that the cumulative effects of these two significant main variables might have on awareness scores. The χ^2 of 4.41 obtained with this group of subjects is significant at the .05 level.

DISCUSSION

The results of this study, in general, give support to the ego-defense hypothesis of conformity. Apparently persons become anxious when they are faced with a unanimous group response at odds with their own. One reaction to this anxiety is to conform, and thus remove the source of the anxiety. It is possible that this is also accompanied by repression of the conflict with resultant unawareness of the original disparity of opinion. Further, it appears that anything which heightens anxiety in the group pressure situation will increase the degree of conformity. That this was a result of the anxiety aroused by the appeal is suggested by the fact that the effect was greater if the person already displayed a high level of anxiety.

The prediction that the effect of heightened anxiety on conformity would be even more pronounced if the person had tendencies to conform as measured by the OAS was not supported. In actuality, it appeared that the persons who tended not to conform on the OAS were most affected by the anticonformity lecture and showed the most conformity under group pressure. The simplest explanation of this phenomenon would be that these people merely refused to conform to the lecturer's demands not to conform. The matter seems not, however, to be that simple. Inspection and statistical comparison of cell means reveal that those subjects low in OAS but high in *MA* scale score reacted with more conformity than did those low in OAS and also low in *MA* scale score. This interaction of anxiety and tendencies to conform is one which should stimulate considerable interest in future research.

REFERENCES

ASCH, S. E. *Social psychology*. Englewood Cliffs, N. J.: Prentice-Hall, 1952.

- BENDIG, A. W., & MCCREARY, J. B. Comparison of two forms of the Manifest Anxiety scale. *Journal of Consulting Psychology*, 1954, 18, 206.
- BLAKE, R. R., & MOUTON, J. S. Conformity, resistance and conversion. In L. A. Berg & B. M. Bass (Eds.), *Conformity and deviation*. New York: Harper, 1961. Pp. 1-37.
- BREGER, L., & RUTZ, C. The role of ego defense in conformity. Paper read at Western Psychological Association, Santa Monica, California, April 1963.
- COUCH, A., & KENISTON, K. Yeasayers and naysayers: Agreeing response set as a personality variable. *Journal of Abnormal and Social Psychology*, 1960, 60, 151-174.
- CRUTCHFIELD, R. S. Conformity and character. *American Psychologist*, 1955, 10, 191-198.
- HOFFMAN, M. L. Conformity as a defense mechanism and a form of resistance to genuine group influence. *Journal of Personality*, 1957, 25, 412-424.
- HOLDER, W. B. Value conformity in normal and non-normal groups. *Journal of Social Psychology*, 1958, 48, 147-154.
- JANIS, I. L. Personality correlates of susceptibility to persuasion. *Journal of Personality*, 1954, 22, 504-518.
- LAWSON, E. D., & STANGER, R. Group pressure, attitude change, and autonomic involvement. *Journal of Social Psychology*, 1957, 45, 299-312.
- MANGAN, G. L., QUARTERMAIN, D., & VAUGHAN, G. M. Taylor MAS and group conformity pressure. *Journal of Abnormal and Social Psychology*, 1960, 61, 146-147.
- SAMELSON, F. Conforming behavior under two conditions of conflict in the cognitive field. *Journal of Abnormal and Social Psychology*, 1957, 55, 181-187.
- SHERIF, M., & HARVEY, O. J. A study in ego functioning: Elimination of stable anchorages in individual and group situations. *Sociometry*, 1952, 15, 272-305.
- SMITH, K. H. Ego strength and perceived competence as conformity variables. *Journal of Abnormal and Social Psychology*, 1961, 62, 169-171.
- TAYLOR, J. A. A personality scale of manifest anxiety. *Journal of Abnormal and Social Psychology*, 1953, 48, 285-290.

(Received March 7, 1966)

RACE, SEX, AND BELIEF AS DETERMINANTS OF FRIENDSHIP ACCEPTANCE¹

CAROLE R. SMITH, LEV WILLIAMS, AND RICHARD H. WILLIS

Washington University

8 samples (total $N = 87$) of white and Negro Ss from northern, border, and southern states rated the friendship acceptance of stimulus persons of specified race, sex, and belief on 1 of 8 general or Negro-white issues. For all samples except 1, belief congruence was more important for acceptance than similarity of race; race, in turn, was more important than similarity of sex. For the southernmost white sample (Louisiana), race was the most important factor by a slight margin, while sex was again least important. Interactions were negligible, except for a renegade effect observed in all 3 Negro samples: members of the racial ingroup were consistently penalized more for disagreeing than were members of the outgroup. None of the samples exhibited "true" discrimination (negative correlation between acceptance of ingroup and outgroup), although the southern white sample exhibited a significant tendency towards true racial discrimination, while all 3 Negro samples showed significant tendencies towards true belief discrimination.

Rokeach, Smith, and Evans (1960) provided evidence supporting a belief-congruence theory of prejudice. According to this theory, similarity of belief about basic issues, rather than racial or ethnic similarity per se, is the primary determinant of social acceptance. When a person rejects a priori a member of another group, it is because he assumes that the outsider holds different views on basic issues. In this way the theory subsumes racial and ethnic prejudice as a special case under belief prejudice. This belief-congruence formulation is quite in the spirit of the structural balance notion of Heider (1958), the ABX model of Newcomb (1959), and the other related theories of cognitive consistency.

In the first of two studies, Rokeach et al. demonstrated that white subjects making

hypothetical choices were more accepting of Negroes who agreed with them on important issues than they were of whites disagreeing with them. This was true both for a northern and a southern sample. In a second study, they showed that Jewish children accepted Gentiles agreeing with them to a greater extent than they did Jews disagreeing with them. Additional support for the theory comes from the research of Byrne and co-workers (Byrne, 1961; Byrne & McGraw, 1964; Byrne & Wong, 1962).

Subsequently Stein, Hardyck, and Smith (1965) demonstrated that both race and belief play significant roles, with race being important in the absence of information about beliefs, but with beliefs on basic issues being more important when both were presented.

In neither study reported by Rokeach et al. were stimulus persons identified as to sex, while in the investigation by Stein et al., male subjects responded to male stimulus persons and female subjects responded to female stimulus persons. All subjects in all three studies were white.

The two investigations reported in this paper represent direct extensions of the "North-South Study" by Rokeach et al. (1960), the one concerning Negro-white relations. In Study I, sex of stimulus persons was introduced as a third determinant of friendship acceptance, along with race and belief.

¹This research is based on two master's theses directed by Richard H. Willis and submitted to the Graduate School, Washington University (Smith, 1966; Williams, 1964). Carole R. Smith's thesis formed the basis for Study I, while Study II is based on that of Lev Williams. We wish to thank the following persons for their much-appreciated assistance in the collection of the data: Vernon L. Allen (University of Wisconsin), C. O. Atchison (Tennessee State University), John Dickerson (Mississippi Vocational College), Vera Kanareff Presbie (Newcomb College), L. Nicholson and T. A. Weir (Harris Teachers College), and Eugene Runyon (Central State University).

²Now at 523 North 18th Street, Columbus, Mississippi.

TABLE 1

DISTRIBUTION OF SUBJECTS IN STUDY I BY SEX AND REGION OF RESIDENCE AT TIME OF STUDY

Sample	Sex		Residence			Total
	Male	Female	North	St. Louis	South	
Wisconsin	23	19	42	—	—	42
Missouri	9	13	—	22	—	22
(white)						
Negro sub-sample ^a	5	16	—	21	—	21
Louisiana	26	29	10	—	45	55

^a The Negro subsample was drawn from the entire Missouri sample.

Two of the three samples contained only white subjects, while the third included a Negro subsample. In Study II, three samples of Negro subjects were used. Both studies incorporated a north-south gradient by the use of one sample from the North, one from a border region, and one from the deep South.

The purpose of the present research is to assess further the validity of the belief-congruence theory of prejudice by (a) pitting belief congruence against both race and sex membership, and (b) utilizing Negro as well as white subjects from various parts of the country.

STUDY I

Except for the extensions indicated, the data-collection procedures and statistical analyses of Rokeach et al. (1960) were followed as closely as possible. Two personality scales, however, were omitted—an anti-Negro scale and Rokeach's Opinionation scale. Both had shown rather low correlations with discrimination responses.

Method

Subjects. Subjects were 140 college students enrolled in an introductory psychology course, distributed by sex and region of residence at the time of the study as shown in Table 1. The Wisconsin sample was recruited from the University of Wisconsin, while the Missouri sample came from Harris Teachers College in St. Louis.³ The Louisiana sample contained subjects from Newcomb College and Tulane University in New Orleans.

Task. Subjects responded to stimulus items consisting of paired descriptions of hypothetical friend-

ship candidates. Pairs of candidates differed regard to race, sex, belief, or combinations of two of these factors. It was the subject's task rate each member of each pair on friendship potential along a 9-point scale. The low end of scale was defined as indicating "I can't see myself being friends with such a person," while the end was defined as indicating "I can very easily myself being friends with such a person."

Experimental design. Subjects were presented a questionnaire containing 192 pairs of stimulus descriptive of hypothetical friendship candidates, stimulus persons. Each stimulus person was assigned a race (white or Negro), a sex, and a belief on one of eight issues. Four of these issues were general nature, while four concerned Negro-white relations. The general issues dealt with belief in God, organized medicine, communism, and labor unions. The Negro-white issues concerned immediate versus gradual desegregation, fundamental equality of race, interracial fraternities and sororities, and freedom for Negroes to own homes wherever they wish.

The stimulus pairs were of six types. With each issue and four pairs of a given type per issue (from combinations of sex and position on the issue), there were 32 pairs of each type. These types were:

- Type R: Difference in race only.
- Type B: Difference in belief only, on a specified issue.
- Type S: Difference in sex only.
- Type RB: Differences in both race and belief.
- Type SB: Differences in both sex and belief.
- Type RS: Differences in both race and sex.

One of the Type R pairs was

- a) A Negro boy who believes in God.
- b) A white boy who believes in God.

while an illustrative pair of Type RB was

- a) A Negro boy who believes in God.
- b) A white boy who is an atheist.

The pairs of all types were presented to subjects in a random order in a mimeographed questionnaire entitled, "Survey of Friendship Choices."

Scoring. Separate scores were computed, for each subject, for the four general beliefs taken together, and for the four Negro-white beliefs taken together. The scores were of two kinds. An *absolute score* was simply the rating from 1 to 9 circled by the subject. A *difference score* was the value obtained by subtracting one from the other of the ratings given to the two members of a stimulus pair.

Signs were attached to difference scores in accordance with information that had been obtained for each subject as to race, sex, and stand on each of the issues. A positive sign indicated a preference in the direction of own race, own sex, or own belief while a negative sign signified a preference in the opposite direction.

³ Although Missouri may possibly be considered a border state, it should be borne in mind that all subjects in the Missouri sample came from St. Louis, and St. Louis is more northern than border so far as relations between the races are concerned.

Results

First we shall consider the 119 white subjects, and later we shall look at the Negro subsample. "The Missouri sample" will hereafter refer specifically to the 22 white subjects from Missouri, while "the Negro subsample" will be used later to refer to the 21 Negro subjects contained in the Missouri group.

Shown in Table 2 are the frequencies with which white subjects preferred *Negroes who agreed* more than, equal to, and less than *whites who disagreed*, for each issue separately. Each subject made two such discrimination responses per issue, one for each sex of stimulus persons in a pair. Because neither

TABLE 2

NUMBER OF WHITE SUBJECTS (STUDY I) WHO PREFER NEGROES WHO AGREE MORE THAN, EQUAL TO, AND LESS THAN WHITES WHO DISAGREE

Belief	Sample	Negro is rated		
		Higher than white	Equal to white	Lower than white
General beliefs God vs. atheism	Wisconsin	63	9	12
	Missouri ^a	37	5	2
	Louisiana	58	8	44
Socialized medicine	Wisconsin	64	9	11
	Missouri ^a	19	13	12
	Louisiana	46	6	58
Communism vs. anticommunism	Wisconsin	84	0	0
	Missouri ^a	40	2	2
	Louisiana	91	6	13
Labor unions	Wisconsin	60	10	14
	Missouri ^a	19	10	15
	Louisiana	45	12	53
Negro-white beliefs Immediate vs. gradual de-segregation	Wisconsin	65	4	15
	Missouri ^a	39	1	4
	Louisiana	52	2	56
Fundamental differences in races	Wisconsin	65	0	19
	Missouri ^a	38	2	4
	Louisiana	55	2	55
Interracial fraternities and sororities	Wisconsin	69	2	13
	Missouri ^a	22	4	18
	Louisiana	31	2	77
Segregated housing	Wisconsin	71	1	12
	Missouri ^a	31	5	8
	Louisiana	52	5	53

Note.—Figures in each row sum to twice the sample size because two entries are tabulated for each subject, one for each sex of stimulus persons in a pair. No noteworthy differences were observed due to sex of stimulus pairs or sex of subjects.
^a White subjects only.

TABLE 3

CORRELATIONS BETWEEN DISCRIMINATION RESPONSES WHEN RACE AND BELIEF ARE VARIED (TYPE RB) WITH RACE DISCRIMINATION (TYPE R) AND BELIEF DISCRIMINATION (TYPE B) RESPONSES: STUDY I

Correlation between race-belief difference scores and	Sample	Negro-white beliefs	General beliefs
Race difference scores	Wisconsin	.19	.21
	Missouri	.18	.21
	Louisiana	.58**	.59**
Belief difference scores	Wisconsin	.91**	.95**
	Missouri	.93**	.96**
	Louisiana	.58*	.45*

* $p < .05$.
** $p < .01$.

sex of subject nor sex of stimulus pairs had any appreciable effect, only the overall frequencies are presented.

For the Wisconsin and Missouri samples, the balance of ratings for each issue was clearly in favor of Negroes who agreed over whites who disagreed. For these two samples, agreement was considerably more important than similarity of race, as predicted by the belief congruence theory of prejudice.

For the Louisiana sample, however, a different pattern was observed. For six of the eight issues, the majority of ratings were in favor of whites who disagreed over Negroes who agreed, although for five of these six issues the differences in frequencies were slight.

For the remaining issue, concerning interracial fraternities and sororities, well over twice as many unequal ratings favored whites who disagreed over Negroes who agreed. Race was somewhat more important than belief in this southernmost white sample, in contradiction to the belief-congruence theory. Rokeach et al., it will be recalled, found belief to outweigh race in both of their groups, including a southern (Houston, Texas) sample.

The same picture is revealed by the correlations in Table 3. If the discrimination responses to Type RB pairs were due solely to race, then such responses should have correlated highly with responses to Type R pairs, but not with Type B pairs. If, on the other hand, responses to Type RB pairs were due solely to belief, they should have correlated highly with Type B pairs, but not with Type

TABLE 4

CORRELATIONS BETWEEN DISCRIMINATION RESPONSES WHEN SEX AND BELIEF ARE VARIED (TYPE SB) WITH SEX DISCRIMINATION (TYPE S) AND BELIEF DISCRIMINATION (TYPE B) RESPONSES: STUDY I

Correlation between sex- belief difference scores and:	Sample	Negro- white beliefs	General beliefs
Sex difference scores	Wisconsin	-.14	-.01
	Missouri	.20	.18
	Louisiana	.01	-.01
Belief difference scores	Wisconsin	.94**	.95**
	Missouri	.94**	.90**
	Louisiana	.93**	.94**

** $p < .01$.

R pairs. The difference between r_{RBB} and r_{RRR} is thus an indicator of the relative influence of the belief and race factors. The correlations in Table 3 show that belief congruence was a much more important consideration than race for the Wisconsin and Missouri samples, but that race was somewhat more important than belief for the Louisiana sample.

Hotelling's test for differences between correlated correlations (Guilford, 1965, pp. 190-191) was used to compare corresponding values of r_{RBB} and r_{RRR} . For the Wisconsin and Missouri samples, all four such pairs of correlations were significantly ($p < .001$) different. For the Louisiana sample, however, neither pair differed significantly.⁴

⁴ Two-tailed tests of significance were employed throughout.

TABLE 5

CORRELATIONS BETWEEN DISCRIMINATION RESPONSES WHEN RACE AND SEX ARE VARIED (TYPE RS) WITH RACE DISCRIMINATION (TYPE R) AND SEX DISCRIMINATION (TYPE S) RESPONSES: STUDY I

Correlation between race- sex difference scores and:	Sample	Negro- white beliefs	General beliefs
Race difference scores	Wisconsin	.94**	.96**
	Missouri	.83**	.93**
	Louisiana	.93**	.95**
Sex difference scores	Wisconsin	.27*	.48**
	Missouri	.37*	.19
	Louisiana	.28*	.23

* $p < .05$.

** $p < .01$.

Fisher's z transformation test (Guilford, 1965, pp. 189-190) was used to test the differences between correlations of the same kind between independent samples. With either kind of belief, no pair of r_{RRR} differed significantly. Within either kind of belief, r_{RBB} for the Louisiana sample was significantly lower ($p < .001$) than that for either of the other samples.

We conclude: race was not significantly more important than belief for the Louisiana sample, but race was significantly more important in this sample—relative to belief—than it was for either of the other samples. Kind of belief, general or Negro-white, made very little difference in any sample.

Correlations analogous to those of Table 4 are used to compare the relative strength of sex and belief (Table 4), and race and sex (Table 5). From Table 4 it can be seen that belief congruence was of overwhelmingly greater potency than was similarity of sex for all three samples. From Table 5, it appears that race was substantially more important than sex, for all three samples. The Hotelling test was applied to appropriate pairs of correlations in both tables, as in Table 3. All differences tested were highly significant ($p < .01$), for all samples, for both kinds of beliefs.

Belief congruence was clearly the most important of the three determinants of friendship acceptance for the Wisconsin and Missouri samples, with race second, and sex a poor third. For the Louisiana sample, race

TABLE 6

DIFFERENCES IN MEAN ACCEPTANCE OF THOSE WHO AGREE AND DISAGREE, NEGROES AND WHITES, AND SAME AND OPPOSITE SEX: STUDY I

Comparison	Sample	Negro- white beliefs	General beliefs
Agreeers vs. disagreeers	Wisconsin	25.61	42.30
	Missouri	28.93	52.28
	Louisiana	11.55	37.71
Negroes vs. whites	Wisconsin	10.51	10.94
	Missouri	15.39	13.40
	Louisiana	53.79	49.75
Same vs. opposite sex	Wisconsin	.31	.82
	Missouri	2.05	-6.37
	Louisiana	4.13	2.84

as the primary consideration, with belief not far behind, and sex again last.

Still another way of demonstrating the same results is through the mean differences in ratings for those who agreed versus those who disagreed, Negroes versus whites, and same versus opposite sex—as in Table 6. The higher mean rating always accompanied the same belief, race, or sex. For the Wisconsin and Missouri samples, the largest differences were due to belief, and the smallest due to sex. For the Louisiana sample, the largest differences were due to race, and the smallest again due to sex. The picture drawn from Tables 2-5 is reconfirmed in all details. It can also be seen from Table 6 that general beliefs were more important than Negro-white beliefs, insofar as they were associated with larger differences between those who agreed and those who disagreed.

Rokeach et al. (1960) argued that *true* racial discrimination implies that the outgroup is discriminated against, but that, at the same time, the ingroup is favored. This definition is in line with Levinson's (1949) conception of ethnocentrism, which implied vilification of the outgroup and overglorification of the ingroup. We should accordingly expect to find a *negative* correlation between acceptance of Negroes and acceptance of whites. Similar arguments could be drawn with regard to those who agree and those who disagree, or with regard to the same and the opposite sex, as ingroups and outgroups.

Instead, positive correlations were found. For the Wisconsin and Missouri samples, acceptance of Negroes and acceptance of whites were substantially correlated, the coefficients ranging between .56 and .77. Thus, for these samples, it is more accurate to speak of misanthropy (or philanthropy) than of discrimination or ethnocentrism. The correlations for the Louisiana sample—although still positive—were quite low, .11 and .20 for Negro-white and general beliefs, respectively. Both were significantly ($p < .01$) lower than the correlations in the other samples, by the z transformation test. Although a general misanthropy dimension was faintly present, since the correlations were not actually negative, the discrimination or ethnocentrism component of

friendship acceptance was relatively prominent.

For Negro-white and general beliefs, respectively, Rokeach et al. obtained correlations of .81 and .79 for the Michigan sample, and correlations of .51 and .35 for the Texas sample. That is, they found slightly less true racial discrimination in the North, and moderately less true racial discrimination in the South, relative to present findings. The difference between the Louisiana and Texas samples was significant ($p < .01$) for the Negro-white beliefs, by the z transformation test. Other differences were not significant.

The correlations between agreeers and disagreeers were also found to be positive, for all three samples and both kinds of beliefs. The range was from .29 to .80, and no north-south gradient was observed. In each case, the correlation for general beliefs was lower than the corresponding coefficient for Negro-white beliefs, but only for the Missouri sample was the difference significant ($p < .01$).

Here again, the Louisiana sample differed from the Texas sample of Rokeach et al. Their sample yielded correlations between agreeers and disagreeers of .20 and $-.14$ for Negro-white and general beliefs, respectively. Their southern sample thus exhibited more true belief discrimination than did the present southern sample, and both differences were highly significant ($p < .002$, $p < .001$).

All three samples showed correlations in the middle and high .90s between acceptance of the two sexes. Acceptance of one sex was almost perfectly correlated with acceptance of the other. Discrimination, in the sense described above, was nonexistent, and individual differences in ratings were due entirely to a general misanthropy (or philanthropy) dimension.

In order to assess the interaction between each pair of variables, mean acceptance ratings were computed for race and belief stimulus subgroups, for belief and sex stimulus subgroups, and for race and sex stimulus subgroups. By and large, interaction effects were negligible. Sex of stimulus persons made little or no difference in the magnitude of preference for agreeers over disagreeers, for example. In addition, sex made very little difference in the extent of preference for whites over

Negroes. For the Wisconsin and Missouri samples, Negroes of the opposite sex were rated lower than Negroes of the same sex, while for the Louisiana sample Negroes and whites of the opposite sex were rated higher than those of the corresponding race of the same sex, but all these differences were inessential.

In the Missouri and Louisiana samples, for general beliefs, there was a slight but noticeable tendency for belief congruence to make more difference in the case of whites, the ingroup, than Negroes. This tendency was not present in the Wisconsin sample nor to any appreciable extent in the Negro-white beliefs for any sample. No conclusion appears warranted. With only occasional and minor exceptions, then, first-order interaction effects were absent.

The Negro subsample. In addition to the 22 white subjects in the complete Missouri sample, there were 21 Negroes. Analysis of the data from this Negro subsample revealed that these subjects reacted very similarly to the white subjects in the Wisconsin and Missouri samples. For all eight issues, whites who agreed were preferred to Negroes who disagreed, by substantial margins, and the size of these margins was comparable to those observed in these two white samples. The belief-congruence theory of prejudice worked just as well for the Missouri Negroes as for the Wisconsin and Missouri whites.

Correlations similar to those presented in Tables 3, 4, and 5 were computed for the Negro subsample, as were mean differences in acceptance like those in Table 6. The general patterning was much the same as that found in the Wisconsin and Missouri samples. Belief congruence clearly had the greatest relative influence, while similarity of sex had the least. Race was still the intermediate factor. Appli-

cation of the Hotelling test showed belief to be significantly ($p < .01$) more important than race for both kinds of persons. Race was significantly ($p < .01$) more important than sex for Negro-white beliefs only. When belief and sex were compared, however, the correlations were significantly different ($p < .01$) only for the general beliefs. As with white subjects, first-order interaction effects were negligible.

The correlations between acceptance of Negroes and whites were .90 for the Negro-white beliefs and the same for the general beliefs. The corresponding correlations for agreeers and disagreeers were .78 and .46, while those for the same and the opposite sex were .99 and .99. These figures are comparable to those for white subjects, except for the correlations between races, which were somewhat higher than the Wisconsin and Missouri samples, and much higher than the Louisiana sample. The difference between the Negro subsample and the Wisconsin sample, for Negro-white beliefs, was significant at the .01 level, while that for general beliefs was significant at the .05 level. Both differences between the Negro subsample and the Louisiana sample were highly significant ($p < .001$). One interpretation of these findings is that these Negro subjects discriminated on a racial basis less than the white subjects—especially the southern white subjects—making a general misanthropy/philanthropy factor sufficient to account for individual differences in level of acceptance.

STUDY II

Method

The only essential difference between Study II and Study I was the nature of the samples. Whereas the large majority of subjects in Study I were white, all 167 subjects in Study II were Negro. These were distributed by sex and region of residence at the

TABLE 7
DISTRIBUTION OF SUBJECTS IN STUDY II BY SEX AND REGION OF RESIDENCE AT TIME OF STUDY

Sample	Sex		Residence						Total
	Male	Female	North	Border	South	East	Midwest	West	
Olema	13	48	27	2	0	23	9	0	61
Tennessee	27	38	0	40	13	5	5	2	65
Mississippi	19	22	1	2	38	0	0	0	41

TABLE 8

Frequency of Responses to Stimulus Pairs
 by Sex, Race, and State, March 1963
 (Total of 1,000 Responses per Pair)

Frequency of Responses

Stimulus	Response	Higher than Negro	Equal to Negro	Lower than Negro
Fundamental Beliefs	Tennessee	28	18	5
	Mississippi	53	21	2
	Ohio	60	13	2
Inter-racial fraternities & sororities	Tennessee	20	17	13
	Mississippi	37	15	10
	Ohio	98	17	7
Segregation in housing	Tennessee	82	22	14
	Mississippi	39	30	13
	Ohio	71	18	13
Fundamental Beliefs	Tennessee	65	14	13
	Mississippi	44	28	18
	Ohio	76	29	17
Inter-racial fraternities & sororities	Tennessee	74	17	19
	Mississippi	37	21	24
	Ohio	82	21	19
Segregation in housing	Tennessee	63	20	17
	Mississippi	44	26	12
	Ohio	97	17	8
Fundamental Beliefs	Tennessee	97	20	13
	Mississippi	58	16	8
	Ohio	97	17	8

Note: Figures in each row sum to 100, as the subjects became two entries are tabulated for each subject, one for each sex of stimulus persons in a pair.

time of the study as shown in Table 7. The Ohio sample was drawn from Central State University in Wilberforce, while the Tennessee sample came from Tennessee Agricultural and Industrial State University in Nashville. Students in both samples were enrolled in psychology classes. The students from the Mississippi sample were recruited from physical education classes at Mississippi Vocational College in Itta Bena. All three institutions are attended primarily or exclusively by Negroes.

Results

Table 8 presents the frequencies with which Negro subjects preferred *whites who agreed* more than, equal to, and less than *Negroes who disagreed*, for each issue separately. Each subject made two such discriminations per

issue, one for each sex of stimulus persons in a pair. For all three samples, and for all eight issues, a substantial majority of subjects preferred whites who agreed to Negroes who disagreed. These data are clearly and consistently in line with the belief-congruence theory of prejudice.

The correlations in Table 9 lead to the same conclusion. The correlations between responses to Type RB pairs and Type R pairs appear to cluster about zero, while those between responses to Type RB pairs and Type B pairs are all high, ranging between .41 and .51. Differences between corresponding RB-B and RB-R correlations were all highly significant ($p < .001$) by the Hotelling test. Again it is demonstrated that belief congruence was more important than race similarity for all three samples.

In contrast to the data from the white subjects in Study I, there was no appreciable overall effect due to the north-south gradient. When the correlations of Table 9 were computed for male and female subjects separately, however, it was noted that the RB-B correlations were consistently lower for females than for males. Furthermore, while differences were small for the Ohio and Tennessee samples, they were appreciable for the Mississippi sample: .93 versus .71 for general beliefs, and .48 versus .51 for Negro-white beliefs. The first difference was significant at the .05 level, and the second at the .01 level by the t transformation test. The female Negro subjects from Mississippi assigned a significantly greater weight to race, relative to

TABLE 9

CORRELATIONS BETWEEN DISCRIMINATION RESPONSES
 WHEN RACE AND BELIEF ARE VARYING (Type RB)
 WHEN RACE IS DISCRIMINATING (Type R) AND
 BELIEF IS DISCRIMINATING (Type B)
 RESPONSES TO STIMULUS II

Stimulus	Response	Negative Correlation	Positive Correlation
Race difference scores	Ohio	.13	.22
	Tennessee	.10	.09
	Mississippi	.02	-.09
Belief difference scores	Ohio	.93**	.90**
	Tennessee	.71**	.71**
	Mississippi	.71**	.82**

** $p < .01$.

TABLE 10

CORRELATIONS BETWEEN DISCRIMINATION RESPONSES WHEN SEX AND BELIEF ARE VARIED (TYPE SB) WITH SEX DISCRIMINATION (TYPE S) AND BELIEF DISCRIMINATION (TYPE B) RESPONSES: STUDY II

Correlation between sex-belief difference scores and:	Sample	Negro-white beliefs	General beliefs
Sex difference scores	Ohio	.26	-.10
	Tennessee	-.01	.10
	Mississippi	.17	-.10
Belief difference scores	Ohio	.83**	.80**
	Tennessee	.58**	.63**
	Mississippi	.55**	.53**

** $p < .01$.

belief, than did their male counterparts. Still, even the females gave most weight to belief.

Table 10 makes use of SB-B and SB-S correlations to show the relative influence of belief congruence and sex of stimulus persons. Correlations between responses to Type SB pairs and Type S pairs appear to cluster about zero, while those between Type SB and Type B pairs are all rather large, ranging from .55 to .83. Belief congruence was more important than similarity of sex in all three samples. Differences were all significant at the .05 level at least, and at the .01 level in most cases, according to the Hotelling test.

The relative influence of race and sex is indicated in Table 11. In all cases but one, the correlation between Type RS pairs and Type R pairs was larger than that between Type RS pairs and Type S pairs. Race was, by and large, more important than sex, but

TABLE 11

CORRELATIONS BETWEEN DISCRIMINATION RESPONSES WHEN RACE AND SEX ARE VARIED (TYPE RS) WITH RACE DISCRIMINATION (TYPE R) AND SEX DISCRIMINATION (TYPE S) RESPONSES: STUDY II

Correlation between race-sex difference scores and:	Sample	Negro-white beliefs	General beliefs
Race difference scores	Ohio	.69**	.10
	Tennessee	.40**	.37**
	Mississippi	.32	.65**
Sex difference scores	Ohio	.41**	.33**
	Tennessee	-.30**	.22
	Mississippi	.06	.22

** $p < .01$.

the two factors were much more evenly matched for the three Negro samples of Study II than for the three white samples of Study I (Table 5). Two out of the three pairs of correlations for Negro-white beliefs were significant ($p < .05$, $p < .01$), while only that for Mississippi was significant ($p < .05$) for the general beliefs.

For all three samples of Study II, then, belief congruence was by far the most important determinant of friendship acceptance. Race was second, and sex again last, although it came in a somewhat better last than it did in the case of the white samples.

A reaffirmation is obtained through the patterning of mean differences in ratings for

TABLE 12

DIFFERENCES IN MEAN ACCEPTANCE OF THOSE WHO AGREE AND DISAGREE, NEGROES AND WHITES, AND SAME AND OPPOSITE SEX: STUDY II

Item	Sample	Negro-white beliefs	General beliefs
Agree vs. disagree	Ohio	47.14	45.99
	Tennessee	45.07	43.66
	Mississippi	33.14	34.73
Negroes vs. whites	Ohio	4.97	2.92
	Tennessee	2.22	1.63
	Mississippi	4.91	6.56
Same vs. opposite sex	Ohio	.53	1.73
	Tennessee	.43	.05
	Mississippi	-.78	1.17

agrees and disagrees, Negroes and whites, and the same and opposite sexes. These appear in Table 12. The belief factor produced the largest differences, while sex produced the smallest, in perfect agreement with the correlational analysis. In addition, a comparison of Tables 6 and 12 reveals that the more nearly even influence of race and sex in Study II was due primarily to the smaller influence of race in Study II, relative to its effect in Study I.

Correlations between acceptance of Negroes and whites, and between males and females, were all high and positive, ranging from .70 to .95. Correlations between responses to agrees and disagrees, on the other hand, ranged between -.09 and .13 and did not differ significantly in any case from zero. True discrimination, in the sense of vilification of

the outgroup and overglorification of the ingroup, was approximated only with regard to belief. With regard to race and sex, a general misanthropy, philanthropy dimension was sufficient to account for individual variations.

The essentially zero correlations between agreeers and disagreeers in Study II stand in sharp contrast with the findings of Study I. The Negro subsample of Study I resembled closely the white samples in this regard. For some reason, the Negro subjects of Study II (but not the Negro subjects of Study I) bordered on indulging in true belief ethnocentrism. Since seven out of eight correlations in Study I were significant, and none even approached significance in Study II, the overall difference between studies can be taken as significant.

First-order interactions were investigated by computing mean ratings of acceptance for belief and race stimulus subgroups, belief and sex stimulus subgroups, and race and sex stimulus subgroups. Interactions between belief and sex, and between race and sex, were quite negligible.

There was, however, a clear-cut interaction between belief and race. For all three samples, and for both kinds of beliefs, mean differences in friendship acceptance of agreeers and disagreeers were larger for Negroes than for whites. This pattern held, without exception, when differences were computed separately for male and female subjects. The two-tailed sign test shows the consistency of this patterning to be significant beyond the .001 level. Furthermore, in 8 out of the 12 instances, Negroes who disagreed were actually rated lower than whites who disagreed. We can say that for the Negro subjects in Study II (but not for those in Study I, recall) there was a tendency to penalize excessively members of their own race who disagreed, as if such people were seen as turncoats or renegades. This *renegade effect*, somewhat surprisingly perhaps, was equally strong for general beliefs as for specifically Negro-white beliefs.

DISCUSSION

The results of both studies, taken together, provide extensive additional substantiation of the belief-congruence theory of prejudice. Not

only most whites, but Negroes as well, allot significantly greater weight to belief than to race in their ratings of friendship acceptance. Furthermore, similarity of sex has been demonstrated to be no competition to belief congruence, or even to race in most instances.

At the same time, the belief-congruence theory sustains one more defeat. The New Orleans whites evidently considered race more important even than belief congruence, unlike the subjects in the Houston sample of Rokach et al. (1960). There is a difference of perhaps 5 or 6 years in the collection of data between the two studies. It is quite possible that there had been during this time an increase in the salience of race as a result of the civil rights movement. If so, this increased salience may have been greater in the South, resulting in a greater relative importance of race.

It is also possible that place, not time, is the operative variable. Houston is some 300 miles west of New Orleans. Moreover, 70% of the students in the Houston sample were born in Texas, a large and culturally heterogeneous state. If substantial numbers of these university students were from western Texas, this could be expected to exert an attenuation of the effect of the institutionalized mores of the deep South. Additional research is needed to clarify the point.

It should be noted that, even if some groups do give most weight to racial and ethnic factors, this is not necessarily fatal for the belief-congruence theory of prejudice. People may often assume, a priori, that members of the outgroup hold basically different beliefs, making belief the underlying consideration after all. Subjects in the Louisiana sample, in particular, may have tended to assume belief incongruence as regards other beliefs in the case of Negroes who agreed. Although this remains to be demonstrated, strong evidence can be cited (e.g., Willis, 1960) that group stereotyping processes need not depend upon motivational factors; cognitive factors can suffice.

A major weakness in the evidence supporting the theory stems from the fact that most

• Additional evidence that the belief-congruence formulation of acceptance works for Negro as well as white subjects is to be found in Stein (1966).

of the directly relevant studies, including the present ones, have employed college students responding to hypothetical and contrived situations. It would not necessarily follow that the theory works equally well in cognitively real settings and/or with less educated subjects. Recent findings, however, indicate that perhaps it often does. Rokeach and Mezei (1966) conducted three experiments on sociometric choice from among real and present persons differing in belief and race. Two of these experiments again used college students, but the third included as subjects job applicants for positions of janitor, laundry worker, and attendant in mental hospitals. In all experiments the choices, which were made following a group discussion, were cognitively real for the subjects. The belief-congruence principle was consistently the best predictor of choices in all three experiments.

It would also appear that the theory is extendable to the area of interpersonal attraction. Several dimensions of similarity in personality have been found to correlate with actual friendship (e.g., Miller, Campbell, Twedt, & O'Connell, 1966) and with marriage choices (Tharp, 1963). More particularly, interpersonal attraction has been found by Newcomb (1961) to be substantially related to agreement on a variety of social issues. The fact that Newcomb's data can be cited in support of the belief-congruence theory, while his theoretical discussion is couched in terms of structural balance, strain towards symmetry, and the like, serves to emphasize the previously mentioned close relationship between the belief-congruence principle and theories of cognitive consistency.⁶

The only noteworthy interaction effect observed was the renegade effect among Negro subjects of Study II. Why do Negroes from de facto segregated colleges, but not those from integrated colleges, nor whites to any extent, penalize members of their own race more for disagreeing than they do members of a racial outgroup? Perhaps this can be

understood as minority-group behavior. Consensus, a concerted pulling together, is more vital to a disadvantaged minority than to a large and relatively secure majority. The dominant group has less need for cohesiveness and can so afford to show more tolerance towards deviant opinion. Certainly Schachter's (1951) finding of greater rejection of deviates in more highly cohesive groups is consistent with this interpretation. A possible point against it, though, is the fact that the renegade effect was fully as pronounced in connection with general beliefs as with Negro-white beliefs.

The lack of interaction between race and sex came as a surprise. It had been anticipated that members of the opposite sex belonging to the other race would be rated as considerably less acceptable as friends than those of the same race, in some samples at least. It was also expected that this effect would be stronger for white than Negro subjects, because of the status differences between the races. These expectations assumed that friendship choices involving the opposite sex would be perceived, not necessarily as implying intimacy, but as implying a definite potential for such intimacy. As interracial sexual intimacy is in violation of the prevailing social norms in many quarters, such implied potential for intimacy might be expected to result in the specific rejection of members of the opposite sex of the other race.

That such an interaction did not occur is probably best explained by assuming that subjects of both races interpreted "being friends with" a member of the opposite sex in a purely platonic fashion. This explanation is made all the more tenable by the finding of Triandis and Davis (1965), especially Table 4 (p. 722), that belief is more important in the case of nonintimate behavioral intentions, while race is more important in the case of intimate behavioral intentions.⁷

Although no sample showed the negative correlations indicative of "true" discrimination in the sense of simultaneous vilification of the outgroup and overglorification of the ingroup, the southern whites bordered on such

⁶ Rokeach and Rothman (1965) have recently extended the belief-congruence principle to the area of cognitive interaction, and in doing so have made more explicit its relationship to another formulation of cognitive consistency, the congruity principle (Osgood & Tannenbaum, 1955).

⁷ See also the exchange between Triandis (1961) and Rokeach (1961).

true racial discrimination, while all three Negro samples in Study II bordered on such true belief discrimination. One rather speculative but provocative conjecture is that de facto segregation encourages true discrimination, the specific type depending upon circumstances. The Wisconsin and Missouri samples, including the Negro subsample, were from integrated educational institutions and exhibited high mean correlations both between acceptance of Negroes and whites, and between agreeers and disagreeers. All three samples in Study II, as well as the Louisiana sample in Study I, were from de facto segregated institutions. The fact that only Negroes from de facto segregated institutions exhibited the renegade effect lends additional plausibility to this conjecture.

REFERENCES

- BYRNE, D. Interpersonal attraction and attitude similarity. *Journal of Abnormal and Social Psychology*, 1961, 62, 713-715.
- BYRNE, D., & MCGRAW, C. Interpersonal attraction towards Negroes. *Human Relations*, 1964, 17, 201-213.
- BYRNE, D., & WONG, T. J. Racial prejudice, interpersonal attraction, and assumed dissimilarity of attitudes. *Journal of Abnormal and Social Psychology*, 1962, 65, 246-253.
- GUILFORD, J. P. *Fundamental statistics in psychology and education*. (4th ed.) New York: McGraw-Hill, 1965.
- HEIDER, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.
- LEVINSON, D. J. An approach to the theory and measurement of ethnocentric ideology. *Journal of Psychology*, 1949, 28, 19-39.
- MILLER, N., CAMPBELL, D. T., TWEDT, H., & O'CONNELL, E. J. Similarity, contrast, and complementarity in friendship choice. *Journal of Personality and Social Psychology*, 1966, 3, 3-12.
- NEWCOMB, T. M. Individual systems of orientation. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 3. New York: McGraw-Hill, 1959. Pp. 384-422.
- NEWCOMB, T. M. *The acquaintance process*. New York: Holt, Rinehart & Winston, 1961.
- OSGOOD, C. E., & TANNENBAUM, P. H. The principle of congruity in the prediction of attitude change. *Psychological Review*, 1955, 62, 42-55.
- ROKEACH, M. Belief versus race as determinants of social distance: Comments on Triandis' paper. *Journal of Abnormal and Social Psychology*, 1961, 62, 187-188.
- ROKEACH, M., & MEZEL, L. Race and shared belief as factors in social choice. *Science*, 1966, 151, 167-172.
- ROKEACH, M., & ROTHMAN, G. The principle of belief congruence and the congruity principle as models of cognitive interaction. *Psychological Review*, 1965, 72, 128-142.
- ROKEACH, M., SMITH, P. W., & EVANS, R. I. Two kinds of prejudice or one? In M. Rokeach (Ed.), *The open and closed mind*. New York: Basic Books, 1960. Pp. 152-168.
- SCHACHTER, S. Deviation, rejection, and communication. *Journal of Abnormal and Social Psychology*, 1951, 46, 190-207.
- SMITH, C. R. Friendship choice as a function of race, sex, and belief. Unpublished master's thesis, Washington University, 1966.
- STEIN, D. D. The influence of belief systems on interpersonal preference. *Psychological Monographs*, 1966, 80 (Whole No. 616).
- STEIN, D. D., HARDYCK, J. A., & SMITH, M. B. Race and belief: An open and shut case. *Journal of Personality and Social Psychology*, 1965, 1, 281-290.
- THARP, R. G. Psychological patterning in marriage. *Psychological Bulletin*, 1963, 60, 97-117.
- TRIANDIS, H. C. A note on Rokeach's theory of prejudice. *Journal of Abnormal and Social Psychology*, 1964, 62, 184-186.
- TRIANDIS, H. C., & DAVIS, E. Race and belief as determinants of behavioral intentions. *Journal of Personality and Social Psychology*, 1965, 2, 715-725.
- WILLIAMS, L. Friendship as a function of race, sex, and belief. Unpublished master's thesis, Washington University, 1964.
- WILLIS, R. H. Stimulus pooling and social perception. *Journal of Abnormal and Social Psychology*, 1960, 60, 365-373.

(Received February 21, 1966)

COGNITIVE AROUSAL:

THE EVOLUTION OF A MODEL¹

GERALD S. BLUM,² JAMES GEHWITZ,³ and CHARLES G. STEWART²

University of Michigan

The heuristic value of a broad conceptual framework for mental processes was explored in the context of the recently popular topic of arousal. The model, pertinent to the possibility for activating "gains" from nonspecific arousal and for differentiating the effects of various sources of arousal upon mental activity. Posthypnotic programming of a trained S was then utilized to induce degrees of anxiety, pleasure, muscle movement, and alertness. A 4th source of arousal, intensities of white noise, was interspersed with the others during a visual discrimination task. The series of experiments began with a general comparative investigation of the influence of these sources upon response latency, followed by more specific probes designed to elaborate the framework.

In our view, one kind of heuristic approach in the development of psychological science begins with a *broad conceptual framework* for mental processes and fills in the details by an orderly progression of laboratory experiments. The present trend in favor of piecemeal theories, concerned exclusively with small segments of behavior, reflects disenchantment with some general formulations of the past. But the goal of evolving an integrative theory capable of linking diverse fields (e.g., perception and psychodynamics) remains desirable and perhaps even essential to a fuller understanding of human thought, feeling, and action. In sketching the outline for a comprehensive model of the mind, one is forced to come to grips systematically with a variety of related issues: complexities beg for analysis and clarification; and innovative techniques for empirical testing are frequently demanded. We shall try to illustrate these taxing virtues in the context of the currently popular concept of arousal (Duffy, 1962; Hebb, 1955; Malmö, 1957, 1959).

Figure 1 is the schematic diagram, developed through discussion of existing theory and research, which will guide our consideration of this topic. (The letter "T" in the

caption refers to alphabetical labeling of successive diagrams dating from 1955.) Sensory input, in addition to activating specific cognitive, affective, or motoric networks, also constitutes a major source of control over amplification in the system. The amount of nonspecific sensory input regulates boosting capacity (level of gain control) of the three variable gain amplifiers, which process signals from their respective networks. The quantity of nonspecific input itself depends, at any given time, upon both external stimulation from the environment and sensory feedback of muscular and glandular responses of the effectors.

Principles governing the formation, activation, and association of networks have been described previously (Blum, 1961). For our present limited purpose only a few background items require mention. The cognitive subsystem early in life begins to develop the capacity to trigger affective and motoric responses. In the mature organism such cognitive mediation is paramount (though the diagram does allow for some sensory input to reach affective and motoric networks more directly).

Access to amplifiers at a particular instant is limited, so that simultaneous cognitive signals compete for amplification. Strong amplified signals set up a feedback loop, and, after reverberation subsides, their networks remain strengthened in memory. Connections are established throughout the subsystem on

¹ The program of research of which this paper is an outgrowth was supported by National Institute of Mental Health Grant 08847-01 (G. S. Blum, principal investigator).

² Now at Stanford University.

³ Now at Tulane University.

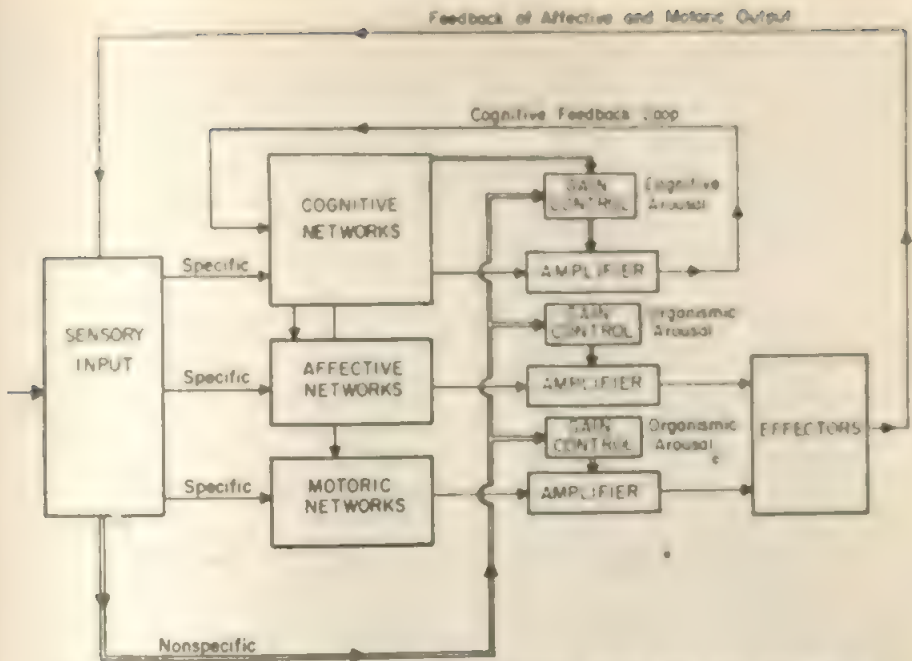


FIG. 1. A model of the mind (Form T). Signal routes (single lines) and amplification control routes (double lines).

the basis of temporal contiguity. One such link connects certain cognitive networks with the gain control of the cognitive amplifier.

The acquisition of cognitive control over amplification serves adaptive functions. In the absence of such control, amplifier overload is more likely to come about. This can happen when exceptionally strong signals from cognitive networks (e.g., "anxiety-laden" thoughts) trigger their linked networks in the affective and motoric subsystems, in turn providing excessive sensory input via effector discharges. Consequent amplifier overload trips a circuit-breaker mechanism as an emergency measure, thereby creating a costly interruption of amplifier activity until automatic reset of the mechanism takes place. Cognitive control offers a more adaptive alternative by lowering the gain of the cognitive amplifier, which forestalls a buildup of very strong signals.

Cognitive control also permits relatively fine adjustment in the amount of amplification available to signals. Situations calling for different degrees of alertness or mental effort can lead to a variety of increased settings of the gain level. Similarly attenuation

can be brought about by signals from diverse cognitive networks. For example, an instruction to "let your mind relax" can be accomplished by lowering the gain level. We have already illustrated the role of attenuation in preventing amplifier overload, where signals from the cognitive representation of anxiety turn down the gain control. Thus selective change in amplification, accompanying signal transmission from specific networks, is a general adaptive property of the system.

Now what are the arousal issues to which this evolving model calls our attention? First we are led to inquire into an apparently meaningful distinction between "cognitive" and "organismic" arousal, based on the assumption that the former reflects gain level of the cognitive amplifier whereas the latter reflects the level of affective and motoric amplifiers. Though a positive correlation between the two aspects of arousal is to be expected (all amplifiers are subject in common to regulation by nonspecific sensory input), the principle of cognitive control over gain highlights a possible discrepancy between them. Translating into the more familiar lan-

guage of psychology, we might question, for example, whether the ubiquitous inverted-U description of the relationship between organismic arousal and performance is uniformly applicable to purely cognitive events.

Previously in some exploratory work with one subject we had noted that vividness of visual imagery tends to be correlated with accompanying galvanic skin responses (GSRs), a common index of organismic arousal. But the relationship is not a necessary one. For example, the provocation of a large GSR by sudden introduction of a shrill whistle had *no* effect upon vividness of the image being visualized in the experimental task at that moment. The subject reported that he had ignored the whistle as something extraneous to the task itself. Cognitive and organismic responses in this instance were totally unrelated. We had also found it rather simple to have subjects, under hypnosis, maintain a constant state of bodily relaxation, measured by basal level of skin resistance (BRL), while mental arousal was varying across a wide range from very low to very high.

Similar observations have not escaped others. Malmö (1963, p. 6) has mentioned some unpublished work from his laboratory which may "pose some difficult questions for activation theory, at least with regard to gauging levels of drive and motivation by means of peripheral physiological indicants." Darrow, Pathman, and Kronenberg (1946) have shown that heart rate and GSR are *negatively* correlated with electroencephalogram (EEG) activation patterns among human subjects who are awake but resting, which suggests that the "cortex and the subcortical structures concerned in arousal can interact antagonistically [Berlyne, 1960, p. 100]." Furthermore, Berlyne asserts that boredom is a state of high arousal caused by cortical inactivation. This specifically implies that the organism as a whole can be aroused at a time when cognitive alertness is low.

A fuller awareness of the cognitive-organismic arousal distinction might clarify certain theoretical disagreements. For instance, Berlyne may have overemphasized the difference between his "high arousal" boredom theory and Hebb's (1949, 1955) "low arousal" formulation. Actually the contrast holds only

for organismic arousal, since both men would apparently agree that monotony produces (and boredom involves) cortical inactivation, that is, low cognitive arousal (Geiwitz, 1966). Fiske (1961) labors over the conflicting evidence from sensory deprivation studies in an effort to determine whether activation is increased or decreased in such situations. In his careful analysis of the issue he does not consider the possibility that arousal may be *both* high and low. That this may be true is suggested by combining Heron's (1957) finding that EEG records during sensory deprivation show deactivation (decreased cognitive arousal?) with the (Vernon, McGill, Gulick, & Candland, 1961) discovery of a BRL drop (increased organismic arousal).

Another issue immediately raised by the model is the comparison of cognitive amplification effects among various *sources* of arousal. The routes in Figure 1 suggest several different sources impinging upon the amplifiers: response feedback, exemplified in affective discharges like anxiety or in purely motoric outputs of muscle movement; external sensory input, such as white noise fed through earphones; and signals from networks in the cognitive subsystem. How might such diverse sources be expected to influence responses to the same cognitive task? What changes would occur if each were manipulated in degree?

All the foregoing questions, posed by initial examination of the model in its crude form, shaped the requirements for the first round of empirical investigation. Hopefully the results would encourage the formulation of more precise questions, in turn resolved by further experimentation, and so on until a definitive account of amplification eventually evolved. Ideally we envisioned a "tracer" which could be sent repeatedly through the cognitive subsystem under various conditions so as to reveal the amounts of amplification actually available for processing. The arousal conditions themselves should be as free as possible of specific cognitive associations, which could obscure the effects in which we were interested. They should also be relatively independent of one another. Each would have to be susceptible to manipulation in degree, evenly maintained in the short run, and re-

liably repeated over long periods of time. Carefully controlled phasing of onsets and offsets would be essential. These manifold requirements are not easily met in conventional laboratory work with human subjects experiencing conditions such as anxiety. Our methodological solutions, which enlisted the aid of hypnotic programming, are described next.

METHOD

In prior years of experimentation we had worked out hypnotic techniques for controlling several affective and other arousal states in degree (Blum, 1963, 1966; Geiwitz, 1966). There was another important legacy—a real, live subject (21-year-old male college student paid to work 4-6 hours per week in the laboratory) who had already participated, along with other subjects, in a series of investigations which covered a 2-year period. We decided to seek answers to our questions by carrying out an intensive *case study by experiment* with this individual, whose prolonged training made him ideally fitted for the exacting controls required to advance the conceptual framework. At this stage of theory building we chose to sacrifice generality for precision in the hope that a model of the mind of Mr. Z would not eventually turn out to be markedly different from others. Some reassurance on this score was provided by the fact that his data in previous experiments had been generally consistent with the rest of the subjects (not all of whom were initially high in hypnotic susceptibility). If we (or, more likely, the reader) need external reinforcement for this decision to concentrate on a single case, we can offer a recent conclusion by Duker (1965, p. 78) from his review of such studies:

Problem-centered research on only one subject may, by clarifying questions, defining variables, and indicating approaches, make substantial contributions to the study of behavior. Besides answering a specific question, it may (Ebbinghaus' work, 1885, being a classic example) provide important groundwork for the theorists.

Manipulation of Arousal States

The subject's existing repertory included the capacity to experience degrees of "free-floating" anxiety, pleasure, and mental alertness in response to certain posthypnotic cues. The hypnotic training procedures are discussed in detail in another publication (Blum, 1966), including a lengthy verbatim account of the steps taken to shape the anxiety responses of this particular subject. Briefly, we begin by having a subject relive, under hypnosis, a relevant situation from his past, for example, one in which he felt considerable anxiety. The feeling itself is then detached by instruction from the original context, manipulated over and over in varying

TABLE 1
POSTHYPNOTIC CUES UTILIZED TO TRIGGER VARIOUS KINDS AND DEGREES OF AROUSAL

Cue	Condition	Degree
100	Anxiety	Very, very strong
70	Anxiety	Fairly strong
40	Anxiety	Some
Large sun	Pleasure	Very, very strong
Medium sun	Pleasure	Fairly strong
Small sun	Pleasure	Some
+MM	Muscle activity	Muscles (below shoulder level) active, twitching, contracting
+M	Muscle activity	Halfway between normal and +MM
-M	Muscle activity	Halfway between normal and -MM
-MM	Muscle activity	Completely relaxed
+AA	Mental arousal	Fever pitch
+A	Mental arousal	Halfway between normal waking level and +AA
-A	Mental arousal	Halfway between normal waking level and -AA
-AA	Mental arousal	Deepest stage of hypnosis or sleep

intensity, and eventually attached to an arbitrary set of cues which automatically trigger responses in the waking state. Table 1 lists the posthypnotic cues employed for the several sources of arousal, including degrees of muscle activity trained specifically for the present studies. During the experiments the subject is simply shown a number (e.g., "70"), a letter cue (e.g., "+MM" or "-AA"), or one of the sun symbols.

When awake, the subject is amnesic for the prior hypnotic programming. In addition to the amnesia, conscious thought associations are removed from the conditions, leaving them nonspecific as far as content is concerned. The subject under anxiety feels that "something bad is going to happen," but does not think about what it might be. Each source is also made as independent as possible of the others. Thus, extremes of muscle activity or mental arousal are trained so as not to create anxiety.

In terms of the model, the above sources sample two types of feedback from affective networks (anxiety and pleasure); internal sensory feedback from purely motoric discharges (muscle activity); and direct, cognitively controlled alterations of amplification (mental arousal). A major remaining source is nonspecific sensory input to the amplifiers from external stimulation. For this purpose four degrees of white noise are fed to the subject through earphones. The decibel levels, along with conventional textbook phrases descriptive of their intensity, are as follows:

- S-4 123 decibels—just below painful sound
 S-3 119 decibels—loud thunder
 S-2 101 decibels—subway train
 S-1 76 decibels—noisy auto

The validity of these arousal manipulations is supported by several kinds of evidence. In the case of anxiety, the subject's introspective report is supplemented by observation of his facial expression and, more convincingly perhaps, by concomitant physiological recording. GSR data dramatically substantiate the induction of degrees of anxiety. Figure 2 gives the magnitude of GSR accompanying all the cue conditions in the course of the first experiment described in the next section, along with that of a neutral or "zero" condition free of any prior hypnotic instruction other than to remain in a "normal" waking state. The conditions are ordered according to the mean millimeters of deflection during approximately a 20-second period. (*N*s for each entry vary between 38–47 trials because of irregular equipment failures.) The three levels of anxiety are clearly differentiated from one another, and the strongest represents an average drop in resistance of approximately 13,000 ohms (1 millimeter = 200 ohms). GSR samples from the same subject in three previous investigations yielded product-moment correlations of .93, .94, and .97 with the four cue intensities of anxiety shown to him (0, 40, 70, 100). An exploratory series of eye photographs showed the same progression of increased pupil diameter with the ascending number cues. In addition to introspective, observational, and physiological support for the validity of these posthypnotic anxiety reactions, the subject's performance in a variety of earlier tasks, involving perceptual recognition, rote memory, and anagram solutions, consistently revealed monotonic effects of the anxiety levels (Blum, 1966).

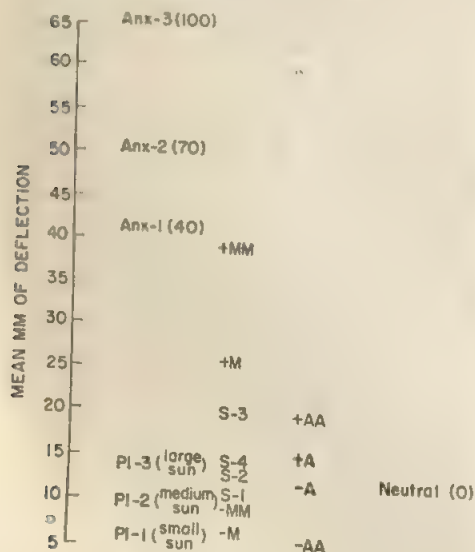


FIG. 2. GSR accompanying various cue conditions (ordered by mean millimeters of deflection).

Our validity information on the pleasure cues is less complete, since there is no commonly accepted independent form of verification, and we have not used pleasure in as many previous studies. Introspective and observational evidence is again unequivocal. The subject's smile, ranging from Mona Lisa to manifest delight, plainly reveals the extent of his free-floating joy. Compared to control trials, the eye photographs consistently show some pupil dilation under the influence of the highest degree, as might be expected from the work of Hess and Polt (1960), but the change in diameter is less than for the low degree of anxiety. The GSR concomitants of pleasure in Figure 2 are negligible, as they have been in other investigations. On the behavioral side, we do know that the three degrees progressively facilitated Mr. Z's performance on a paired-associate learning task.⁴ Also his subjective feeling of certainty with respect to perceptual judgment is spuriously high in the case of very strong pleasure, which suggests the desired presence of euphoria.

Mental arousal, like anxiety, can be gauged directly from the subject's facial expression. During mental activity he reports changes in ability to concentrate when the different cues are presented—stronger concentration going with higher arousal. A test of his skill in performing a visual tracking task yielded a remarkable set of ordered, virtually non-overlapping curves for the various mental arousal conditions (Ehrlich, 1964). Physiologically, GSR effects in Figure 2 are slight, though again ordered from low to high cues, and pupil size does not differ from controls.

The last two manipulations do not require much additional confirmation. The GSR data document the magnitude of muscular response, but any observer can immediately distinguish the overt muscle twitches of the two strong cues from each other and from the low ones. The white-noise condition only makes use of hypnosis to remove from consciousness any specific associations to the sounds and to insure that the subject does not become anxious at the highest levels. Figure 2 indicates fulfillment of the latter requirement, for none of the sound intensities produces mean deflections in the anxiety range.

In summary, hypnosis is utilized as a tool to assist in relatively pure and precise mobilization of the several sources of arousal implicated by the conceptual framework. Prevailing controversies over the nature of hypnosis are not immediately pertinent, our primary concern being to demonstrate the validity of manipulations capable of replication by others.⁵

⁴ Nancy Reynolds, "Hypnotically Induced Levels of Pleasure in Paired-Associate Learning," unpublished manuscript, University of Michigan, 1964.

⁵ For an excellent review of hypnotic phenomena, skeptical attitudes, and studies of individual susceptibility, the reader is referred to Hilgard (1965). A tentative theory of hypnotic compliance suggested by our model is also available (Blum, 1961).

A Task to Measure Amplification Available in the Cognitive Subsystem

The remaining technical decision was to select a task which would provide the "tracer" to be sent through the cognitive subsystem, under the various arousal conditions, as a test for the amount of amplification available with each. The stimuli we chose were a series of seven 1-inch-square photographs of an X on a white background, differing only in darkness of line. They were created by developing a photograph of a single line drawing for varying durations and can be described as a graduated scale progressing by fairly equal subjective intervals from a blank to a very black X. These were presented to the subject in a Gerbrands tachistoscope at an exposure period of .036 second—a speed picked after extensive pretesting because it made the task of intermediate difficulty and did not show practice effects after several hundred trials.

The subject's ostensible task was to identify which of the seven possible Xs was flashed on a given trial, and also to rate his certainty on a 4-point scale immediately afterward. They were labeled "A" through "G" with accompanying code words like "bell" for "B" in order for his verbalization to be clearly understood. The X series was shown tachistoscopically once before the start of each day's session, along with the correct identifications, but accuracy feedback was not provided after individual trials during the experiments. In reality the major dependent variable in our design was response latency, which we considered to be the most sensitive indicator of possible arousal effects. If more amplification is available during the perceptual and cognitive phases of forming a judgment, then processing should be accelerated. Stronger signals are presumed to eventuate in a response more quickly.

An incidental though not unimportant methodological advantage of the latency measure is that the subject never became aware of the fact that his responses were being timed. A Standard timer (.01-second resolution), activated by the flash of the tachistoscope and stopped by the subject's verbal report by means of a transistorized voice-operated relay, was concealed in an outer room to which he did not have access. The microphone in the experimental room had been a standard piece of equipment on the scene throughout his participation in earlier studies. This point has relevance for the hypnotic subject's presumed susceptibility to demand characteristics of the situation. The subject's ignorance of the latency measure was accompanied by the fact that the outside recorder did not know to which conditions the subject was responding inside the experimental room. Other precautions routinely used in our research include instructions to the subject at the very outset of his work in the laboratory that the best opportunity to serve the experimenters' purposes is just to "let things happen naturally," followed by carefully conducted waking and hypnotic inquiries at the conclusion of an experiment.

Besides recording response latencies, the operator in the outer room also monitored continuous chart

recording of GSR. The latter apparatus included a Fels dermohmmeter, modified by a 16 cycle per second "chopper" to reverse polarity and an Edin pen recorder. Zinc finger electrodes were used.⁶

Thus equipped with techniques for the desired manipulations of arousal plus a task designed to reflect changes in available amplification, we proceeded with our attempt to chart the mind of Mr. Z.

EMPIRICAL STUDIES

Comparative Effects of Arousal Conditions

The initial experiment consisted of an extensive exploration of effects of the five sources of arousal, each controlled in degree, upon response latency in the identification of flashed Xs by the subject in the waking state. The experimenter began a trial by inserting a cue into the adapting field of the tachistoscope. The subject, always wearing earphones with a constant (66 decibels) background white noise, immediately reacted to the cue with the appropriate experience. After 10 seconds the cue was removed, serving as a ready signal that the flash was soon to come. The subject, however, continued responding to the cue in accord with prior hypnotic instruction. Following a variable delay (randomly assigned) of from 3 to 9 seconds, the stimulus was then flashed. The subject's cue reaction was programmed to last *through* the instant of the flash and to dissipate immediately afterward. The latter instruction was intended to keep the purely motoric aspects of verbalization relatively free of cue influence, thus confining the experimental effects more to the perceptual-cognitive realm.

Sound trials began with a "vertical sine wave" cue (merely an indication of a sound trial, calling for no response on the part of the subject) which was immediately removed. The white noise was then generated in 5-second cycles at the prescribed intensity and the X flashed after the number of cycles covered 18 seconds plus the randomly assigned

⁶ We wish to acknowledge the technical assistance of Joseph Horner, who devised and built the chopper circuit, assembled the timing apparatus, and generally maintained the equipment throughout these and other experiments. Jackson Beatty, Angeline Ehrlich, Ruth Millar, and Carol Reeves performed as operators in the outer room and also assisted in subsequent data analyses. Daniel Kahneman aided by Jackson Beatty and William Baker carried out the pupil photography described earlier.

delay period. It should be noted that on sound trials the flash occurred *after* the experimental condition had ceased.

There were 19 independent variables (4 levels of muscular response, 3 of anxiety, 3 of pleasure, 4 of white noise, 4 of mental arousal, and the neutral or "zero" condition). The overall design consisted of 60 trials for each of the 18 arousal variables (10 pairings of the cue with each of 6 Xs, omitting the blank) and 270 trials in the zero condition, making a grand total of 1,350 trials. The experiment filled 15 2-hour sessions of 90 trials each, within which the pairing of cues and X stimuli was randomized.

Table 2 presents the mean response latencies for the several degrees of each of the five sources along with the neutral condition. The overall analysis of variance yielded significance for treatment conditions well beyond the .01 level. In order to provide a distribution-based measure for rough estimation of relative effects, the overall mean and average standard error are included in the table. Median latencies, arranged in order, are plotted in Figure 3.

All degrees of anxiety slow down processing, the magnitude of effect increasing monotonically with strength of anxiety. Pleasure tends to produce short latencies in its three degrees. The strongest, Pl-3, has the fastest median of all 19 conditions (Figure 3), but the

TABLE 2

EXPERIMENTAL EFFECTS OF VARIOUS TYPES AND DEGREES OF AROUSAL UPON RESPONSE LATENCY

Condition	Mean latency*	Condition	Mean latency
Anxiety		Pleasure	
Anx-3 (100)	653	Pl-3	431
Anx-2 (70)	607	Pl-2	456
Anx-1 (40)	535	Pl-1	435
Muscle activity		White noise	
+MM	643	S-4	411
+M	590	S-3	359
-M	450	S-2	371
-MM	460	S-1	391
Mental arousal		Neutral	
+AA	435	Zero	431
+A	479		
-A	521		
-AA	598		

Note.—N = 60 except Neutral where N = 270.

* Mean latencies in hundredths of a second, that is, 100 = 1 second. Overall mean of distribution = 493, average standard error = 29.

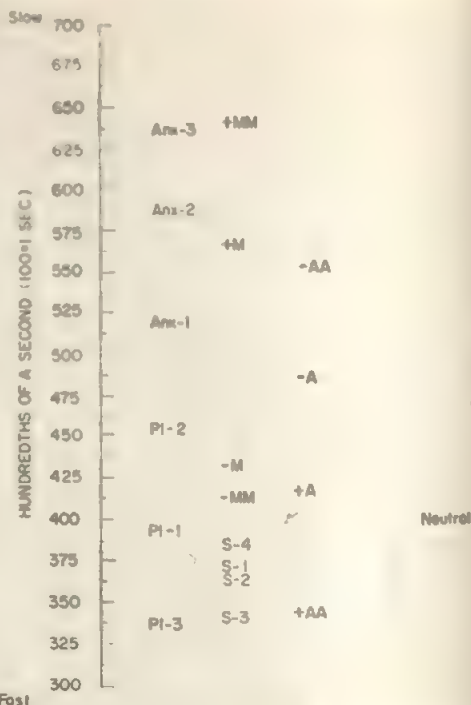


FIG. 3. Order of median latencies for cue conditions.

greater skewness of its distribution produces a mean latency no different from neutral. Strong muscle activity has slowing effects comparable to anxiety, +MM close to Anx-3 and +M to Anx-2; the two conditions of muscular relaxation (-M and -MM) are slightly slower than neutral. Latencies for the levels of white noise are all short, below the neutral point, and the differences among them are small. The mental arousal cues, on the other hand, have remarkably monotonic effects, spanning a broad range of latencies. The fastest, +AA, corresponds closely to Pl-3 and S-3 in the median order, whereas the slowest, -AA, approximates +M and Anx-2. The intermediate levels of over (+A) and under (-A) arousal fall between the extremes, though the former is somewhat slower than neutral. The disparity of mean and median for +AA is again indicative of greater skewness, as in the case of Pl-3.

The supplementary data on accuracy and certainty parallel those of latency. Significantly more errors (compared to neutral) are made at Anx-2 and Anx-3, +M and +MM, -A and -AA. All pleasure and sound levels,

+A, and +AA result in somewhat fewer errors than neutral. The direction of X calls is of interest in this connection. An index relating each condition to the neutral pattern of calls reveals that all the major slowing cues in Figure 3 are accompanied by *undercalling*, the order of effect from most to least being -AA, Anx-3, Anx-2, -A, +MM, +M, and Anx-1. In other words, these cues lead the subject to report the flashed X at a lighter position on the scale than he does in the neutral condition. The index does not point to any clear overall evidence, though all four sound levels exceed neutral.

The subject's certainty ratings also conform closely to the latency data. The three levels of anxiety, the two of overactive muscle movement, and the two of low mental arousal all yield significantly lower certainty than neutral, each set in expected order. Pl-3 and +AA are both significantly higher, with the nine remaining cue conditions bunched near neutral.

The concomitant GSR data, in addition to their previously discussed role as validator of the arousal component of hypnotically induced conditions, bear on our conceptual distinction between "cognitive" and "organismic" arousal since skin resistance is a prime index of the latter. A glance back at Figure 2 suggests that the larger deflections (anxiety cues and +MM) are associated with slowing down of response latency, whereas more moderate drops in resistance (+M, S-3, +AA) vary in their effects from slow to fast. The absence of appreciable deflection (Pl-1, -AA, -M) is also ambiguous in its relationship to latency, since only -AA is correlated with slow responding. Taken as a whole, the GSR data at best provide only marginal support for the inverted-U function of arousal commonly described in the literature.

A closer look reveals more convincingly that the experimental effects cannot be accounted for solely in terms of concomitant autonomic arousal. A mean latency analysis by cue condition was carried out on all trials where GSR exceeded 20 millimeters of deflection. Though *N*s in each of the 19 conditions varied considerably in this subset of the data, resulting latencies duplicated the ordering of Table 2 almost exactly. The sub-

set of trials at the other extreme (0-5 millimeters of deflection) permitted many fewer latency comparisons among cues, but several of the original effects were still discernible. Thus, with organismic arousal essentially held constant (statistically), the various cue conditions continue to exert their differential influence upon response latency.

Referring to the model at this point, we have valuable leads concerning differential consequences of various arousal routes for the amount of amplification available for processing whatever happens to be introduced, in the present case a tachistoscopically flashed X. A number of questions vital to detailed elaboration of the conceptual framework suggested themselves next. Where in the system are the effects occurring? Can they be localized in the beginning of the sequence at the time of perceptual registration in the cognitive subsystem, or subsequently in the process linking the perceived X to the memory file of the X scale? Are there specific effects beyond the cognitive subsystem, upon the verbal response per se, despite the instruction for the cue reaction to terminate prior to verbal report? Can the action of individual cues be broken down as far as the condition itself is concerned? For example, what are the relative contributions of organismic versus cognitive components of anxiety reactions to the observed decrease in available amplification? Also, how might different sources act in combination, that is, two concurrent negative influences or two positive ones? All these questions provided impetus for the following rounds of experimentation.

Localizing the Obtained Effects

In order to discover the "locus," an experiment was carried out to measure the influence of the cues on two phases not easily separable by conventional methods: perceptual registration or "perceiving" in a more restricted sense, and cognitive judgment or the comparison of perceived stimulus with the file of Xs in memory. The procedure, somewhat similar to that employed in the first experiment but involving selected cues and a smaller *N*, was identical in form for both the perceiving and the judging parts. The subject began by responding to a cue. When the X flashed, he

closed his eyes and "fixed the image" of what he had just seen. In other words, he took a sort of mental photograph, without making any decision as to which X it was. Indeed, due to prior hypnotic instruction, his mental file of Xs was not consciously available to him during this process. When he had fixed the image clearly in his mind, he was to say "now" (producing a measurable latency by means of the voice relay), put the image aside for later recall, and open his eyes. A new cue was then inserted. After 9 seconds this cue was removed and the subject closed his eyes. Six seconds later he heard a distinct click through the earphones, which was a signal to "bring back" the previously fixed image along with his now-available file of Xs and to judge which X it was (latency again recorded).

In the perceiving phase, eight cues (+AA, -AA, +MM, Anx-3, Pl-3, S-3, S-4, and zero) were randomly assigned to the first phase of this two-part procedure, the image fixing. After viewing the stimulus under the influence of one of these conditions, the image was brought back by the subject and judged under the zero or neutral condition. In the judging phase of each trial, image fixing was done under the neutral condition, whereas judging was affected by one of the eight cues. Thus the influence of each condition on perceiving and judging, separately, was assessed by response latency.

Another deviation from the original procedure was that the effect of the cue (other than sounds) had been hypnotically pro-

grammed to last through the response, for a fixed period of time, in both parts of the sequence. In order to control for cue influence on responding per se, a set of simple response trials was interspersed throughout the experimental ones. In the perceiving phase such trials were preceded by a card saying "no image," and the subject merely observed the flash of the blank X stimulus ("A"), closed his eyes without fixing an image, and said "now." In the judging phase the advance signal card contained a prearranged X response, for example, "D," and later at the click the subject just repeated the prescribed X label instead of forming a judgment.

Table 3 presents the latency data for the two phases. Pure response effects were first partialled out by pairing experimental and control trials and adjusting the mean experimental latencies by covariance analysis. Cue differences achieved overall statistical significance in the perceiving phase ($p < .05$) but not in the judging ($p < .25$). The directions of effect in the perceiving phase parallel those of the earlier study: high anxiety, lowered mental arousal, and to a lesser extent strong muscle activity continue to produce decremental effects; Pl-3 and +AA both accelerate responding, as do the loud sound levels to a smaller degree.

The finding that arousal conditions were exerting their effects primarily upon the registration phase of processing led us into two additional avenues of empirical exploration. A series of 48 trials with the same eight cues provided ratings from the subject, for each flash, on scales of clarity, darkness, and uniformity (arms of X bent, etc.). Xs seen under Anx-3, -AA, and +MM were consistently reported as more blurred, lighter, and less uniform. Identical results were obtained in another series of 48 "slow motion" trials. With the X remaining in full view for 1 minute on each trial, the subject gave a running commentary of how it looked to him—the first 15 seconds of the trial under the neutral condition, the next 30 seconds under one of the eight cues, and the last 15 seconds back under neutral. Also, in the course of hypnotic inquiry conducted after completion of the experiments, the subject reported that he was unable to differentiate the appearance of Xs

TABLE 3
EXPERIMENTAL EFFECTS OF AROUSAL CONDITIONS
UPON PERCEIVING AND JUDGING

Condition	Perceiving—mean latency ^a	Judging—mean latency ^b
+MM	657	663
-AA	682	653
Anx-3	721	706
Neutral	636	725
Pl-3	559	692
+AA	507	613
S-3	595	616
S-4	585	567

Note.—N = 24 in each condition.

^a Mean latencies in hundredths of a second, adjusted to partial out response effects. Overall mean of distribution = 618, average standard error = 43.

^b Mean latencies in hundredths of a second, adjusted to partial out response effects. Overall mean of distribution = 654, average standard error = 43.

TABLE 4
EXPERIMENTAL EFFECTS OF AROUSAL CONDITIONS
UPON RESPONDING PER SE

Condition	Mean latency ^a
+MM	308
-AA	277
Anx-3	270
Neutral	254
PI-3	223
+AA	222
S-3	199
S-4	197

Note.—*N* = 60 in each condition.

^a Mean latencies in hundredths of a second. Overall mean of distribution = 244, average standard error = 24.

in the Anx-3, -AA, and +MM conditions, all of which made it seem as if he were looking through a "smeared glass or a dirty window on a camera." From the limited data available to us from photographs of the pupil under Anx-3 and -AA, this blurring phenomenon common to both cues does not seem to be due to dilatation changes because lowered mental arousal did not alter pupil size. By implication the phenomenon is more likely cognitive than organismic.

Turning to the question of arousal effects upon responding per se, we have already noted that the subject's reactions in the first experiment were hypnotically programmed to cease immediately after the flash. During later hypnotic inquiry the subject confirmed that the cue experiences had worn off prior to his verbalization, with the exception of the overactive muscle responses which tended to subside more slowly and often lasted up to his report (an interesting admission from the standpoint of demand characteristics, since this was counter to the hypnotic instruction). However, we were curious about effects on verbalization itself and carried out a series of experiments with trials similar to the response control ones in the perceiving versus judging study, that is, where the subject knew in advance what was to be flashed and no judgment was involved. These response control data were combined statistically to produce an overall distribution of cue effects upon response, shown in Table 4. The results indicate a strong slowing effect for +MM, possibly due to muscular competition and interference as the subject attempted to vocalize in the presence of widespread twitching and jerking.

The other conditions also reveal directions of influence similar to those in the first experiment. Additional response control series, in which (like the first experiment) the cue was programmed *not* to act through the responses, did not yield any appreciable differences in latency among conditions, which supports the interpretation that latencies in the first study cannot be attributed to cue effects upon verbalization.

Breaking Anxiety Down into Its Cognitive and Organismic Components

Another question emerging from the initial investigation concerned the relative influence of two aspects of the anxiety reaction: the ever-present impression, not quite expressible in words, that "something bad is going to happen" versus the physical sensations of general body tenseness and stomach disturbance, all of which characterized in varying degree his responses to the number cues. In terms of the model, the distinction is between the activation of a significant cognitive network and feedback from effector discharge which serves as potent nonspecific input to the amplifiers. How does each of these components contribute to the obtained effect of progressively slower processing with higher degrees of anxiety?

For this study the subject was trained to respond with the anxiety "thought only" or the "bodily feeling only" or both thought and feeling intact as before. The design of the first experiment was repeated with randomly alternated sequences of trials in each condition, including neutral and some PI-3 trials interspersed for relief. The experiment was run over 4 sessions and included 42 trials in each of the 9 anxiety conditions; 126 neutral, and 42 PI-3 trials. The results are plotted in Figure 4. (As an aside, it should be mentioned that concomitant GSR recordings validate the posthypnotic anxiety responses—the mean deflections in the "bodily feeling only" conditions being virtually identical with those in the intact conditions, whereas the "thought only" conditions all produced slightly smaller mean deflections than did neutral.)

From the first column in Figure 4, where mean latencies for the intact cues are plotted,

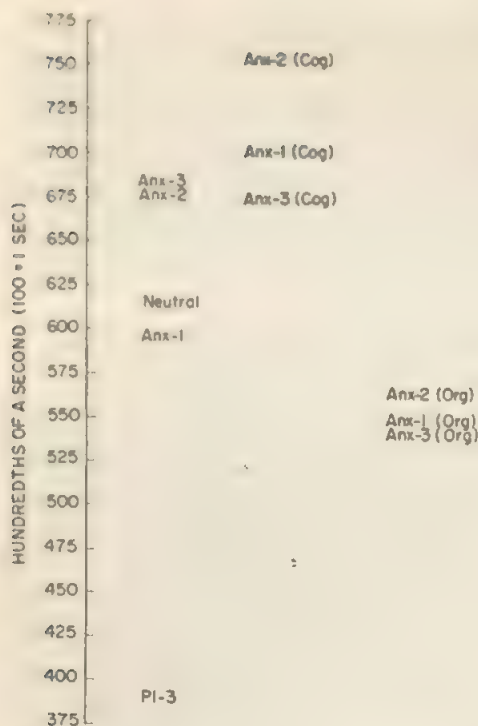


FIG. 4. Mean response latencies for various anxiety components, neutral, and strongest pleasure.

we see that the original order of findings is replicated in four out of five instances, the exception being neutral which is slower than in Figure 3 and appears on the other side of Anx-1. The second column, the cognitive (Cog) "thought only" condition, shows Anx-3 essentially unchanged in its position, but both Anx-2 and Anx-1 have an even greater decremental influence. The third column, the organismic (Org) "bodily feeling only" condition, has the three cues closely grouped slightly below the neutral point—no evidence at all for a decrease in available amplification.

Combined Action of Arousal Conditions

The last two sections have described follow-up studies which sought to break down the experimental effects in more detail, by localizing them along the perceiving-judging-responding continuum and, in the case of anxiety, by analyzing a complex arousal reaction into its component parts. Another kind of question important for the evolution of our model involves the action of different

conditions in combination rather than singly. Here we chose for intensive investigation the two conditions which had exerted a facilitating influence in the original experiment, very strong pleasure and heightened mental arousal. Again the same design was employed except that the subject was programmed to respond simultaneously to the two cues shown together in the adapting field of the tachistoscope before the X flash. There were six combinations in all, PI-3 and no pleasure being paired with +AA, +A, and a "normal" level of mental arousal instructed to be halfway between +A and -A. Forty-two randomized trials were run for each of these combinations.

The mean response latencies, given in Table 5, clearly indicate the *additive* nature of the interaction. The shortest latencies occur for the PI-3 and +AA combination, the longest for the neutral pairing, and the intermediate in ordered progression for the remaining combinations. The generality of an additive interpretation is supported by a very limited exploration ($N = 6$ in each cell) of two decremental conditions acting in concert: Anx-3, Anx-2, and no anxiety, each paired with +MM, +M, and "normal" muscle activity. Despite the small N , the latencies again conformed strikingly to the additive model.

DISCUSSION

At this point the attempt to chart amplification routes in the mind of Mr. Z had to be abandoned short of completion. Just as the next series of questions and empirical tests began to take shape, the United States Navy stepped in to claim his services. However, we believe that the data already in hand permit us to conceptualize the influence of diverse forms of arousal upon mental processes in

TABLE 5
MEAN RESPONSE LATENCIES FOR VARIOUS COMBINATIONS OF PLEASURE AND MENTAL AROUSAL

Pleasure	Mental Arousal		
	+AA	+A	Normal
PI-3	368	442	484
None	514	611	622

Note.— $N = 24$ in each cell.

more detail than has heretofore been possible. In addition to further theoretical questions, there remains, of course, the necessity to extend the findings to other subjects and different cognitive tasks. Within these limits, what can we now say about cognitive arousal?

Our efforts focused on that aspect of cognitive arousal designated as *available amplification* for processing whatever inputs happen to be active in the cognitive subsystem. The experiments reveal notable congruence among the effects of strong anxiety (Anx-3), lowered mental arousal (-AA), and overactive muscle movements (+MM). All are characterized by very long response latency, poor accuracy in the direction of undercalling, low subjective certainty, and description of the Xs as "fuzzier and fainter." There are differences in degree of influence, and +MM appears to contain more of a response effect, but by and large the similarities are striking. The amount of available amplification seems to be reduced in each case.

Hypnotic inquiry revealed that for all three it was "hard to concentrate," but they differed in their cognitive accompaniments: Under anxiety the subject was preoccupied by the premonition that "something bad was going to happen," which he "felt but didn't think about in words"; under -AA he could not concentrate because he felt sleepy and "didn't think about anything at all"; and under +MM he "couldn't think as well," but didn't think about anything else outside the task. In the latter condition he also complained that involuntary jerking of his head (although hypnotic instruction applied only to the body below shoulder level) made it difficult for him to see the X clearly.

At the facilitating end, heightened mental arousal (+AA), very strong pleasure (Pl-3), and the four white-noise levels appear to increase available amplification. The subject reported that both +AA and Pl-3 made him feel alert, although his state of high motivation in +AA involved watching "to see what would happen," whereas in Pl-3 it went along with feeling happy. ("You feel good and awake and more aware of everything around you.") He felt that the two lower sound levels did not make him more alert and the higher ones bothered him somewhat.

The evidence strongly suggests that these diverse effects are not tied to organismic aspects of the arousal conditions. Directly pertinent is the investigation of anxiety components, where the decremental results were associated with cognitive but not organismic responses. Also the initial experiment revealed that essentially similar X phenomena were accompanied by very high and very low autonomic activity (Anx-3 versus -AA in Figure 2); equally low GSR scores (-M and -AA) differed considerably in their concomitant X effects.

It might be argued that some types of organismic arousal, not reflected by drops in skin resistance or increases in pupil size, could be related more closely to the empirical results. The very loud white-noise stimuli are a possible case in point, since 123 decibels administered to both ears can reasonably be expected to exert organismic influence despite its minor GSR activity. Though S-3 and S-4 do accelerate responses to the flashed X, the two lower levels, S-1 and S-2, are no less effective. The interpretation we favor is that all four levels served cognitively as an extra ready signal for the flash inasmuch as the X appeared after the sounds subsided, a clue which Mr. Z reported afterward. As a check we ran a lengthy series of trials in which the sounds were delivered concurrently with the flash (not done originally because of a decision to keep the viewing context uniform across conditions). The results were clearly reversed, all sound levels being associated with *longer* latencies. A very few exploratory trials with some of the other cues, programmed to end 1 second before instead of just after the flash, hinted that their usual influence might disappear. These observations underline the significance of the timing of cue activity for future study, but they also tend to minimize the role of organismic factors in the task. Split-second timing considerations should not be so crucial for their operation.

At this juncture one might question whether the task itself is at all susceptible to organismic influence. Some incidental evidence from the first experiment yields a clearly affirmative answer. The relatively large number of neutral trials (18) on each of the 15 days required to complete the data collection per-

mitted the analysis of daily fluctuations in latency. On some days the neutral trials were faster or slower than on others, with mean latency differences ranging from $\frac{1}{2}$ to 1 full second on either side of the overall mean (cue conditions, it should be noted, were counter-balanced over days). BRL scores for neutral trials turned out to be significantly lower (increased organismic arousal) on the fast days and higher (decreased arousal) on the slow, compared to the medium days. In terms of the model, the normal operating level of gain in the several amplifiers, based on the subject's degree of general alertness or sleepiness on a particular day, does have a direct bearing on the speed of cognitive processing. But the various arousal conditions, sometimes producing very large alterations in autonomic response during short intervals, seem to be having their experimental effects primarily within the cognitive subsystem. Temporary organismic responses therefore must be differentiated, within the conceptual framework, from those more enduring in nature.

Turning now to the cognitive factors involved in the arousal conditions, we can conceptualize the facilitating effects as a consequence of momentary increases in gain level of the cognitive amplifier (see Figure 1). This boosting is accomplished by signals transmitted to gain control from various networks inside the cognitive subsystem: heightened mental arousal conditions, strong nonspecific pleasure ("more aware of everything around you"), and sound levels interpreted as an extra alerting phenomenon. Furthermore, results of the combination experiment, involving degrees of heightened mental arousal and pleasure, point to an additive model for two such sources acting together.

Conversely, decremental effects in the case of decreased mental arousal can be attributed to direct lowering of the gain control on the cognitive amplifier. The pervasive cognitive repercussions of this attenuation are manifested not only in the radically altered registration of the X, but also by the inability to sustain attention to the task.

Interpretation of the role of strong anxiety is not so easy. We know that its effects are very similar to those of -AA, in distorting perceptual registration and impairing concen-

tration, so the assumption of a common decrease in available amplification is tenable. But there are several ways in which anxiety can be conceived to lessen the amount for use in processing the X. One, the temporary breakdown of amplifier activity in response to nonspecific sensory overload, is contraindicated by the anxiety components experiment in which the "bodily feeling only" condition, where such overload might also have been expected, had no appreciable influence.

A second possibility involves division of attention. The cognitive component of anxiety, though vague and not expressible in words, nevertheless could be competing successfully with the X for amplification—the stronger the anxiety cue, the less opportunity for the X to be processed through the cognitive amplifier. This explanation is not supported by the subject's ratings of the arousal conditions according to the amount of his attention claimed by each during the time the X was being flashed. The same high rating, that is, greater distractibility, was assigned to +AA and Anx-3, which had opposite experimental effects; the same intermediate rating was given to Pl-3 and Anx-1, again very different in their influence; and a low rating to +MM is also at odds with the obtained data. (As an aside, we can report that subsequent work with another subject in a new cognitive task demonstrates that anxiety continues to produce comparable decrements when the cognitive component is removed from consciousness and allowed to operate only at a preconscious level.)

Another distinct possibility, based on prior model formulations of the inhibitory role of anxiety (Blum, 1961), pictures the cognitive component of anxiety as lowering the gain on the cognitive amplifier and slowing the X response in the process. Presumably this link between cognitive anxiety signals and lowered gain starts to form early in life and gradually gets strengthened, by repeated contiguity in the course of anxiety discharges of sufficient magnitude to break down amplifier operation temporarily. Subsequently any input, such as the "100" cue, which activates the cognitive component of anxiety will lead to the transmission of an attenuation signal and a consequent decrease in amplification.

This function, not unlike Freud's (1936) "warning signal" conception of anxiety, is adaptive because it weakens cognitive signals en route to affective and motoric networks and forestalls potentially more disruptive effector outputs. A parallel interpretation can be applied to excessive muscular discharge, as in the +MM condition, by substituting a cognitive representation of muscle overload for the anxiety component.

A definite choice among these alternative ways of conceptualizing the role of anxiety awaits further research. The distraction view, unlike the other two, does not predict attenuation of the anxiety response itself. However, the present data do not permit this differentiation because the anxiety reaction was deliberately maintained at a constant level by instruction.

In a sense, this has been a progress report on the distance our Model T has traveled along the road to better understanding of cognitive arousal. Taking previous theory and research as a point of departure, the broad conceptual model has carried us through post-hypnotic exploration of several avenues to arousal: degrees of anxiety, pleasure, muscle movement, cognitively controlled alertness, and white noise. Each route was mapped empirically with the aid of a visual discrimination task, yielding latency of the subject's responses in identifying tachistoscopically flashed Xs of varying darkness.

REFERENCES

- BERLYNE, D. E. *Conflict, arousal, and curiosity*. New York: McGraw-Hill, 1960.
- BLUM, G. S. *A model of the mind*. New York: Wiley, 1961.
- BLUM, G. S. Programming people to simulate machines. In S. S. Tomkins & S. Messick (Eds.), *Computer simulation of personality*. New York: Wiley, 1963. Pp. 127-157.
- BLUM, G. S. Hypnosis in psychodynamic research. In J. E. Gordon (Ed.), *A handbook of clinical and experimental hypnosis*. New York: Macmillan, 1966. Pp. 83-109.
- DARROW, C. W., PATHMAN, J., & KRONENBERG, G. Level of autonomic activity and electroencephalogram. *Journal of Experimental Psychology*, 1946, 36, 355-365.
- DUFFY, E. *Activation and behavior*. New York: Wiley, 1962.
- DUKES, W. F. $N = 1$. *Psychological Bulletin*, 1965, 64, 74-79.
- EHRLICH, N. J. Hypnotic control of a compensatory tracking task. *Perceptual and Motor Skills*, 1964, 19, 232-234.
- FISKE, D. W. Effects of monotonous and restricted stimulation. In D. W. Fiske & S. R. Maddi (Eds.), *Functions of varied experience*. Homewood, Ill.: Dorsey, 1961. Pp. 106-145.
- FREUD, S. *The problem of anxiety*. New York: Norton, 1936.
- GEIOWITZ, P. J. Structure of boredom. *Journal of Personality and Social Psychology*, 1966, 3, 592-600.
- HEBB, D. O. *The organization of behavior*. New York: Wiley, 1949.
- HEBB, D. O. Drives and the C.N.S. (conceptual nervous system). *Psychological Review*, 1955, 62, 243-254.
- HERON, W. The pathology of boredom. *Scientific American*, 1957, 196(1), 52-56.
- HESS, E. H., & POLT, J. M. Pupil size as related to interest value of visual stimuli. *Science*, 1960, 132, 349-350.
- HILGARD, E. R. *Hypnotic susceptibility*. New York: Harcourt, Brace & World, 1965.
- MALMO, R. B. Anxiety and behavioral arousal. *Psychological Review*, 1957, 64, 276-287.
- MALMO, R. B. Activation: A neurophysiological dimension. *Psychological Review*, 1959, 66, 367-386.
- MALMO, R. B. On central and autonomic nervous system mechanisms in conditioning, learning and performance. *Canadian Journal of Psychology*, 1963, 17, 1-36.
- VERNON, J. A., MCGILL, T. E., GULICK, W. L., & CANDLAND, D. K. The effect of human isolation upon some perceptual and motor skills. In P. Solomon et al. (Eds.), *Sensory deprivation*. Cambridge, Harvard University Press, 1961. Pp. 41-58.

(Received March 2, 1966)

CONCEPTUAL LEVEL AS A COMPOSITION VARIABLE IN SMALL-GROUP DECISION MAKING¹

PAUL STAGER²

Princeton University

Decision making was investigated from the standpoint of the emergent distribution of functional roles, conflict generation, utilization of conflict in decision synthesis, and information acquisition. 20 4-man groups of male undergraduate and graduate students were composed to yield 4 combinations of conceptual level. Each group participated in a complex simulated tactical decision-making situation. The group interaction was coded into functional role categories and rated with respect to conflict generation and utilization. The information measure *H* was used to assess uncertainty in the functional role distributions, in terms of which group structure was construed. Uncertainty of group structure, utilization of generated conflict, and search for novel information were linearly related, while interpersonal conflict was curvilinearly related to an increasing percentage of members of a high conceptual level in the group.

Small groups active in the decision-making process are concerned with the three distinct but interrelated functions of information acquisition, information processing, and subsequent decision making. The effectiveness of the decision-making process in the groups would seem to depend upon the quality of the differentiated component processes and the degree to which the three functions are integrated within the groups.

There is considerable evidence in the small-group literature (Bass, 1960; Collins & Guetzkow, 1964; Steiner, 1964) that heterogeneity of group composition may in some instances facilitate and, in others, be detrimental to the integration of group processes. The level of performance on certain tasks is frequently a function of the extent to which group processes are integrated. The effect of group composition on performance is dependent upon the type of situation or task confronting the groups and upon the composition variables. There is a definite interaction effect between personality and situational variables. Both sets of variables act as intervening variables in mediating human performance

and have been viewed as variable "complexes" (Ware, 1964).

One critical personality and composition variable which has been shown to be a determinant of performance in decision-making tasks is the conceptual level dimension (Schroder, Driver, & Streufert, 1967); this dimension describes the integrative complexity of a group member's conceptual structure. The term conceptual structure refers to a set of cognitive mediating links which produce a relatively stable group of techniques by which the individual receives, processes, and transmits information. The integrative complexity of the conceptual structure is a function of the number of dimensions along which stimuli are ordered, the number of different schemata with which the perceived dimensions of information are organized, and the complexity of the organization. Individuals whose information processing is characterized by the use of few dimensions of information and few or fixed schemata in a given domain are described as having a *low conceptual level*; individuals who typically perceive many dimensions of information and utilize many alternate combinatory schemata are described as having a *high conceptual level*. Although the level of information processing tends to increase with an increase in the level of conceptual structure, the former can vary as a function of various forms of environmental stress.

¹Based in part on a doctoral dissertation submitted to the graduate faculty, Princeton University. Reported in part at the 1966 meeting of the Eastern Psychological Association, New York.

This study was supported by the Office of Naval Research under Contract Nonr 1858-(42).

²Now at York University.

The previous studies which have used conceptual level as a composition variable (Stager & Kennedy, 1965; Tuckman, 1964, 1966) focused primarily upon group performance rather than group processes. Since there is an interaction effect between personality or composition variables and situational variables, it becomes necessary to examine differences in group processes when considering performance differences. Tuckman (1966) has attempted to relate the composition variable, heterogeneity, to group performance. The manner in which composition heterogeneity affects group processes, however, remains to be clarified. If certain task situations require decision-making groups to structure themselves in certain ways, then it is necessary to know what type of group information-processing structures will emerge in given group compositions.

The information-processing interpretation of personality presented by Schroder et al. (1967) suggests that some individuals are more readily able to cope with environmental uncertainty than others. The interpretation also suggests that the way in which decisions or judgments are made and the predecisional subprocesses which are emphasized are functions of the way in which an individual processes information. In a group situation, the predecisional subprocesses would seemingly be differentially accommodated by the group members, depending upon their characteristic manner of processing information. Using this type of interpretation, prediction of emergent structuring, given certain group compositions, becomes possible.

If effective decision making is dependent upon the range of alternatives or diversity which the group generates in the predecision phase, the conceptual level of the group members also has particular relevance in predicting the nature of the predecisional processes. Schroder et al. have theorized that the generation of conflict and diversity becomes increasingly prevalent in information processing as the conceptual level of the individual increases. Evaluation and integration of discrepant information, which are integral subprocesses of decision making, are also theorized to be more characteristic of information processing at a high conceptual level.

Finally, effective decisions cannot be made without an adequate (or optimal) amount of information about the decision situation. Recent research (Stager & Kennedy, 1965; Suedfeld & Streufert, 1966) has indicated that the nature of the information-search activity and the type of information sought are functions of the conceptual level of the individual or group members. Different compositions of group members, composed on the basis of the conceptual level dimension, would be expected, therefore, to exhibit differences in their search for information.

As previously suggested, there is the necessity of understanding the nature of the characteristic group structures and the concomitant group processes, given certain group compositions, in explaining performance differences. The present study was directed toward differences in group functioning in an attempt to relate the conceptual level dimension to the emergent group information-processing structures and the characteristic predecisional subprocesses. Four specific hypotheses, derived from the conceptual systems theory (Schroder et al., 1967), are presented in the following paragraphs.

Hypotheses

Hypothesis 1. With an increase in the percentage of members of a high conceptual level in the group, there is an increase in the role flexibility or, conversely, a decrease in structuring and more functional role uncertainty.

Theoretically, high conceptual level members are able to cope with a higher level of uncertainty in their environment and are more adaptable to environmental demands than low conceptual level members. Whereas high conceptual level members may be aware of considerable uncertainty in a situation, without attempting to increase the amount of structuring (reduce uncertainty—Sieber & Lanzetta, 1964), low conceptual level members tend to simplify and structure their environment; for example, groups homogeneously composed of low conceptual level members form hierarchical group structures (Tuckman, 1964). In groups comprised of members with different levels of conceptual structure (heterogeneous groups), clearly defined boundaries between subgroups or parts

are likely to develop. If differentiation occurs on the basis of functional roles or task boundaries, heterogeneous groups should be characterized by a high degree of functional role centrality (Hutte, 1965), particularly when there is a low percentage of members of a high conceptual level in the group. The uncertainty in the functional role structure (the amount of uncertainty in the group information-processing structure), therefore, was predicted to increase as the percentage of high conceptual level members in the group increased.

Hypothesis II. Groups in which the members are all of a high conceptual level (100%) generate more interpersonal (substantive) conflict than groups in which the members differ in conceptual level.

Individuals of a high conceptual level characteristically generate diversity and conflict in their processing of information (Schroder et al., 1967). For high conceptual level group members, "... reality is defined as being possessed of multiple alternatives and hence diversity is sought as a means of enhancing validation [Schroder & Harvey, 1963, p. 157]." Similarly, Bennis and Shepard (1956) have noted that in groups which are advanced in their development conflict is undisturbingly present; conflict is generated from a delineation of substantive issues.

Hypothesis III. The extent to which generated conflict is utilized in the synthesis of decisions increases with an increasing percentage of members of a high conceptual level in the group. With an increasing percentage of high conceptual level members, there is increasingly more synthesizing of generated alternatives and evaluating of alternatives in the predecisional phase.

The second and third hypotheses are interdependent since the capacity to synthesize alternatives is concomitant with the capacity to cope with diversity and conflict in the generation of alternatives. Although low conceptual level members may generate numerous essentially unrelated alternatives, they are not able to evaluate them as extensively as the high members. Since the former are less sensitive to the extent to which diversity is being generated, in groups with low percentages of members of a high conceptual level diversity

and conflict are less likely to be maintained at optimal levels. In other words, there is likely to be a superoptimal level of diversity with little integration or evaluation of discrepant information. When groups are functioning at a higher conceptual level, they act as effective integrative instruments (Martin & Hill, 1957), in which there is utilization of differences and collaboration among members (Hearn, 1955). Schroder and Harvey (1963) have stated that group functioning, at a high conceptual level, is characterized by a consensus which is the result of rational discussion rather than a compulsive attempt at unanimity.

In groups comprised of predominantly low conceptual level members, the development of a hierarchical group structure prevents fluidity of the functional role structure and minimizes the possibility of there being conflicting and divergent alternatives generated, or at least evaluated. Moreover, since low conceptual level members, unlike high level members, are not characterized by the ability to make multiple discriminations or to assume different perspectives in regard to give discrepant units of information, the prediction was made that evaluation and synthesis would characterize the participation of high conceptual level members.

Hypothesis IV. The extent of search for novel information increases as the percentage of members of a high conceptual level in the group increases, whereas total information search is not dependent upon group composition.

Since high conceptual level members are able to cope with more uncertainty, to generate additional information from that already in memory storage, and to utilize a more complex dimensional structure in the perception of the environment (Schroder et al., 1967), the assumption was made that groups with a higher percentage of high conceptual level members would exhibit less extensive searching for other than specifically novel information.

Lanzetta (1963) has explored the relationship between environmental uncertainty and information-search behavior, suggesting that information search is elicited by a response conflict engendered by response uncertainty:

the greater the degree of uncertainty, the greater is the conflict and the stronger the motivation to search. Intuitively, it would seem that individuals seek information until they reach some optimal level of uncertainty, at which they make a decision, the level being a function of the level of initial uncertainty, the rate of uncertainty reduction, and the level of conceptual structure. The inability of low conceptual level members to cope with higher levels of uncertainty would tend to require more extensive information search before decisions could be made.

Stager and Kennedy (1965) found that information search was inversely related to an increasing percentage of high conceptual level members in the group when there were no imposed explicit costs for information search, but directly related to an increasing percentage of high conceptual level members with imposed explicit costs. A differential decrease in the level of the total information search (groups with a greater percentage of highs exhibited little change in the level of information search, whereas groups with a lower percentage of highs exhibited a marked decrease), in the latter condition, suggested that there was a critical amount of information necessary for different levels of conceptual structure. A finding (Suedfeld & Streufert, 1966), which is possibly related to the observed differential decrease, is that the proportion of novel to total information requests was significantly higher for high conceptual level individuals than low conceptual level individuals; the latter, while making a higher proportion of information-search moves, were primarily concerned with receiving informational feedback about ongoing activities. In contrast, the high level individuals searched for information about new, previously unexplored aspects of the situation. The finding provides a basis for an interpretation of the effects of imposed search costs; the lesser decrease in the level of search activity, by groups comprising a higher percentage of high conceptual level members, may be taken to indicate that the groups with a lower percentage of highs decreased their level of search by omitting information about relatively less critical aspects of the environment when search costs were imposed. The groups

with a higher percentage of high conceptual level members, on the other hand, were primarily concerned with only the novel or essential aspects under both cost conditions.

METHOD

Subjects

Subjects were 80 male university undergraduate and graduate students. The subjects were selected from among approximately 500 volunteers on the basis of the measures subsequently described.

Subject Selection

The subjects were administered the following tests:

1. The Paragraph Completion Test of Conceptual Level (Schroder et al., 1967). The Paragraph Completion Test is a projective test of several sentence stems designed for the assessment of the level of conceptual structure. On the basis of previously acquired norms, subjects were classified as low in the level of conceptual structure if they obtained scores of 3 or less, and high in the level of conceptual structure if they obtained scores which were greater than 7.

2. The n Dominance scale of the Edwards Personal Preference Schedule (Edwards, 1959). The need for dominance items reflects the desire to be a leader, to give advice, to make decisions, and to defend one's own position. For purposes of comparison the scale was administered in the same form as used by Turkman (1966), in order to reduce the visibility of the scale, it was given with the n Affiliation scale of the Edwards Personal Preference Schedule, and the items from the two scales were interspersed. n Dominance scores were converted to quartile scores, based on norms derived from a sample of approximately 200 individuals; subjects were classified as high on n Dominance if they scored in the highest quartile, intermediate if they scored in either of the two intermediate quartiles, and low if they scored in the lowest quartile. This breakdown was used since the research design required half of the subjects to be intermediate on this measure, with half of the remaining subjects high and the other half low.

Group Composition

Those subjects classified as high or low in the level of conceptual structure were considered for further selection; all others were rejected. There was no difference in the levels of intelligence, as measured by the Wonderlic Personnel Test, between the high and low conceptual level subjects. The subjects were further subdivided into high, intermediate, and low n Dominance groupings. Twenty four-man groups were composed to yield four combinations of conceptual level. Each of the four combinations was represented by five groups. The group compositions were defined with respect to the dimension of an increasing percentage of members of a high conceptual level in the group. The homogeneous com-

position comprised four members who were of an equally high conceptual level (100%). The other types of composition comprised groups in which three, two, or one of the members, in each type of composition, respectively, had a high level of conceptual structure. The groups, therefore, were considered as being equally spaced along this dimension. The 100% group was homogeneous in composition and the 25, 50, and 75% groups were heterogeneous with respect to the conceptual level dimension. Member *n* Dominance was controlled by systematically varying the distribution of high, intermediate, and low *n* Dominance members across members of each group composition. Each group comprised one high, one low, and two intermediate *n* Dominance members. In each instance, at least one of the high conceptual level members in the group was of intermediate *n* Dominance. Groups were matched, as closely as possible, on intelligence.

Experimental Decision Environment

Each group participated in a complex simulated tactical decision-making situation (Streufert, Clardy, Driver, Karlins, Schroder, & Suedfeld, 1965). The groups were confronted with a model of an island which they were to assume was held by an enemy force of unknown strength and location. Acting as four members of equal status on a military field staff, they were instructed to engage the enemy and to secure the island. The groups received information about the enemy movements and the effects of their own decisions by providing for acquisition of such information through the deployment of their own forces. Responses to their decisions were provided by a preprogrammed input (Karlins, Schroder, & Streufert, 1965), which was perceived by the groups as realistically dynamic and responsive. The fixed-input program was initially designed for the purpose of providing a controlled and standardized input of information to subjects participating in the tactical situation. The duration of the tactical decision situation comprised seven ½-hour periods.

Coding and Rating Scales

Preceding the experiment, a coding scheme was derived and found, after continuous refinement, to yield reliable functional categorization of individual and group behavior. The final coding and rating scheme was prepared in the form of a manual which could be used by the observers.

Verbal behavior of the groups was coded according to the predefined categories of perceiving and proposing the problem, requesting information, supplying information, suggesting alternatives, evaluating alternatives, autocratically deciding, and confirming decisions through consensus. Each category was considered as a functional role in the decision-making process; scoring, therefore, was concerned with the changes of a group member from one role to another. In order to provide additional analyses, the frequencies with which (a) new or novel information search was requested, (b) different alternatives were

proposed, or (c) evaluations of different alternatives were given, while members were in the respective functional roles, were noted.

Groups were rated with respect to their utilization or synthesizing of conflicting alternatives, the degree to which interpersonal conflict was present, and the number of effective communication channels which were available to the group members (i.e., communication complexity).

The reliability of coding and rating, across all of the measures, was determined by computing the Pearson product-moment correlation coefficient of the independent assessments. The reliabilities, based on the assessments of two trained observers, ranged between .79 and .96 for the first 11 experimental groups. The median reliability coefficient, .91, was taken as justification for the assumption that observations by a single observer would be reliable for the remaining 9 groups. Although both observers were aware of the theoretical assumptions, the fact that one observer was not informed of the group composition and the rapidity with which the coding had to be done were assumed to minimize confounding between assessments and the group composition variables.

Assessment of Group Structure

Two methods were used to assess the structuring which emerged in the different group compositions. *Group information-processing structure* was construed in terms of the amount of uncertainty, existing within the group, with respect to the frequency with which the different group members tended to perform various decision-making functions (i.e., assume different functional roles). The group information-processing structure, construed in this manner, was tractable by the information measure of uncertainty (H). If it is assumed that the different functional roles represent different categories of events, then by noting the frequency with which the different functional roles are assumed, a frequency or probability distribution for the categories can be generated. The uncertainty involved in the distribution is obtained by applying the Shannon measure of average information (see Garner, 1962). The formula for computing the average information is as follows:

$$H = - \sum_i p_i \log p_i$$

In the formula, i represents any one of the individual alternatives (functional roles) available. Maximum uncertainty for any distribution occurs when all categories are associated with equal probabilities. In reference to the total group uncertainty, the uncertainty or H value for each member's participation can be determined and then summed with the H values for the other members, thus providing a total H value for the group. A higher value of H for a group would be obtained when the members tend to exhibit "flexibility" by assuming different roles at different times in sequential decision making. Groups, alternatively, could reduce the amount of

certainty in their information-processing structure, interpersonal environment, by organizing into a decision-making structure in which the different members assumed different functional roles consistently. It is important to note, at this point, that the uncertainty involved in a member's participation resides not in the transmitter (the member concerned), but in the receivers (the remaining group members and the external observer); the computed uncertainty for a given member's participation may correspond to his subjective uncertainty. The latter assumption is not critical, however, in the present conceptualization.

The first measure applied to the data concerning group participation was, therefore, the Shannon measure of uncertainty. The second measure was the role-centrality index which has previously been described by Hutte (1965). The centrality index is a value, between 1 and 0, which indicates the extent to which only one member assumes a given functional role (indicated by an index value of 1), or all of the members assume a given functional role equally often (indicated by an index value of 0). The degree of role centrality is determined by applying the formula,

$$C = \frac{(t-1)d_h - (d_{h-1} + d_{h-2} + d_{h-3})}{(t-1)D},$$

where t is the total number of group members, D is the total number of contributions in the decision category (or functional role) concerned, d_h is the number of contributions by the most active member (i.e., the highest score), and d_{h-1} represents the second highest score. For each group, the centrality indexes for each functional role were summed across the different roles, thus yielding a *total structural index*.

Procedure

Each of the 20 four-man groups participated in the tactical decision-making situation with the assumption that the enemy was represented by another group. The group interaction, generated by the task, was tape-recorded and observed through one-way observation screens. The members' participation was recorded by means of individual throat microphones connected to separate tape recorders. During the 3½-hour session, the group interaction was coded continuously according to the described categories or functional roles. Groups were rated, at the end of each ½-hour session, on the behaviors previously listed. Each of the orders drafted and submitted by the groups was retained for further analysis.

RESULTS

Group Structuring

Figure 1 presents the mean values of the total group uncertainty H for each type of group composition. An analysis of variance indicated that the composition effect was sig-

nificant ($F = 12.9$, $df = 3, 16$, $p < .001$); the increase in H , with an increasing percentage of members of a high conceptual level, was significantly linear (with a trend analysis yielding $F = 37.3$, $df = 1, 16$, $p < .001$). The trend analysis applied to the data assumed equal distances between the different compositions. The assumption is defensible since the different compositions were derived by replacing one additional low conceptual level member with one high conceptual level member at successive intervals along the group-composition dimension. Figure 1 also presents the mean values of the total structural index. As illustrated, the uncertainty measure H and the structural index appear to provide comparable assessments of the degree of group structuring. The composition effects were significant ($F = 4.5$, $df = 3, 16$, $p < .05$), the increase in group structuring (as measured by the structural index), with an increasing percentage of high conceptual level members also being significantly linear ($F = 17.6$, $df = 1, 16$, $p < .001$). Total group structuring, as reflected in the total structural index, correlated significantly ($p < .01$, $df = 18$, one-tailed test) with group uncertainty H ($r = -.62$). The latter correlation provides a more accurate representation of the relationship between group uncertainty and the total structural index than Figure 1 itself. The points plotted in Figure 1 are, as indicated, based on means of five groups each and are intentionally scaled to correspond as closely as possible, in order to emphasize the comparable trends which were observed in the data.

Interpersonal Conflict and Conflict Utilization

Composition means derived from ratings of interpersonal (substantive) conflict and utilization of conflict in the synthesis of decisions are presented in Figure 2. The degree of interpersonal conflict present in the group interaction was curvilinearly related to group composition (a quadratic trend analysis yielding a significant $F = 13.0$, $df = 1, 16$, $p < .01$); the rated utilization of generated conflict increased linearly (linear trend analysis $F = 66.3$, $df = 1, 16$, $p < .001$) with an increase in the percentage of highs in the group. An analysis of variance indicated that

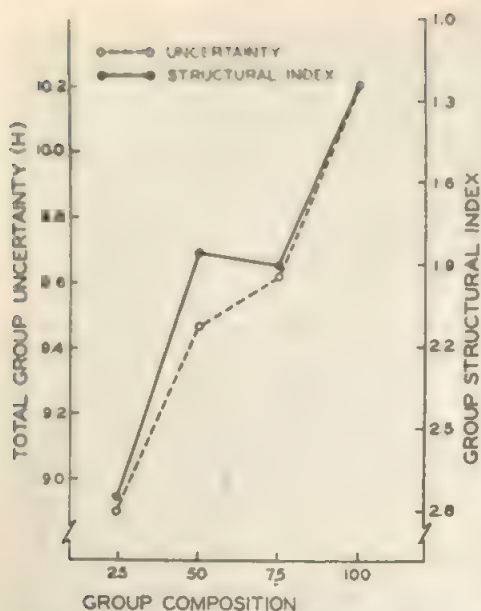


FIG. 1. Group uncertainty (H) and structural index as functions of an increasing percentage of members of a high conceptual level in the group.

the composition effects for interpersonal conflict were significant ($F = 4.7$, $df = 3, 16$, $p < .05$); the homogeneous (100% highs)

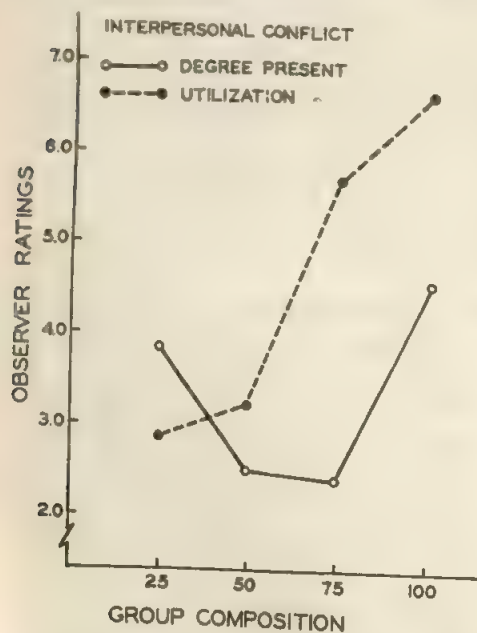


FIG. 2. Interpersonal conflict and conflict utilization as functions of an increasing percentage of members of a high conceptual level in the group.

groups generated significantly more conflict ($p < .05$ in a post hoc comparison of means) than any of the other compositions. Similarly, composition effects, with respect to the utilization of generated conflict were significant ($F = 23.8$, $df = 3, 16$, $p < .01$).

Generation of suggestions or alternatives did not vary significantly with different group compositions; the number of evaluations, however, and hence, the ratio of suggestions to evaluations (S/E), was subject to composition effects (Figure 3). Analysis of variance yielded a significant F in both instances (evaluations, $F = 3.4$, $df = 3, 16$, $p < .05$; S/E ratio, $F = 15.9$, $df = 3, 16$, $p < .01$). A trend analysis indicated that the S/E ratio involved both linear and quadratic components ($F_s = 135.0, 8.3$, $df = 1, 16$, $p < .01$, $p < .025$, respectively); the number of evaluations increased linearly ($F = 9.9$, $df = 1, 16$, $p < .01$).

Figure 3 also illustrates the linear increase in the complexity of group communication (linear trend $F = 10.1$, $df = 1, 16$, $p < .01$) that occurred as the proportion of highs in the group increased. Rated communication complexity or, alternatively, openness of com-

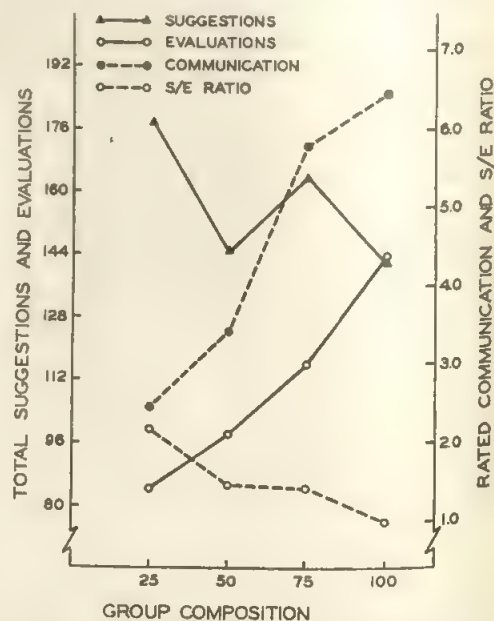


FIG. 3. The effect of an increasing percentage of members of a high conceptual level in the group on the generation of alternatives, evaluation of alternatives, communication complexity, and S/E ratio.

communication channels, correlated significantly ($p < .01$, $df = 18$) with the rated utilization of generated conflict ($r = .95$), the S/E ratio ($r = -.67$), and the number of evaluations made by the group ($r = .53$). Total group structuring (structural index) also correlated significantly ($p < .01$, $df = 18$, one-tailed test) with the S/E ratio ($r = .59$) and the rated utilization of conflict ($r = -.63$). The partial correlation coefficient for group structure (structural index) and the S/E ratio, with the effect of communication removed, was not significant. A variance interpretation of the coefficients indicated that some 86% of the association of the latter three variables resulted from the effect of communication. Communication ratings correlated significantly with the total group structural index ($r = -.70$) and group uncertainty ($r = .73$).

Information Search

As in the case of the group information-processing structure, two measures, one derived from an analysis at the group level and one from an analysis at the individual level, were used to assess group search for specifically novel information (Figure 4). A direct assessment was obtained by determining the number of information-search orders which were actually submitted by the groups. As illustrated in Figure 4, the number of orders increased in a linear manner (linear trend $F = 14.0$, $df = 1, 16$, $p < .01$) across the various compositions. Composition effects were significant ($F = 5.41$, $df = 3, 16$, $p < .01$).

The second assessment was obtained by combining the frequencies with which different members in a group requested novel information; the assessment included proposals by the members for the group to take some kind of action which would provide novel information. The second assessment, therefore, indicated the propensity within the group for information search, rather than the actual execution of information-search orders. Again, composition effects were significant ($F = 3.46$, $df = 3, 16$, $p < .05$). Although the combined frequencies increased linearly (trend analysis $F = 7.4$, $df = 1, 16$, $p < .025$) with an increase in the percentage of high conceptual level members, a comparison of

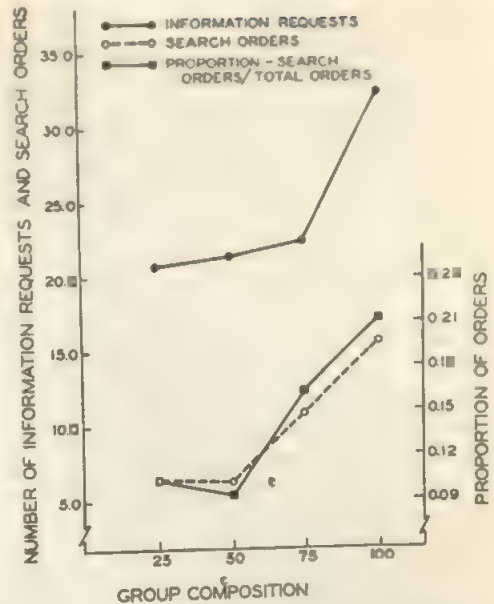


FIG. 4. Information requests, search orders, and proportion of search orders for novel information as functions of an increasing percentage of members of a high conceptual level in the group.

the group means indicated that the significance was attributable to the difference between the mean of the 100% high groups and the average of the means for the other compositions (Figure 4). The relationship of the number of information-search orders (novel information), expressed as a proportion of the total number of orders submitted, is illustrated in Figure 4.

For purposes of assessing the total information search in the groups, the frequencies, with which the functional role of requesting information was assumed, were summed across the group members; no significant differences between the various compositions were obtained.

DISCUSSION

The results confirmed the hypothesis (I) that group structuring and role differentiation decrease with an increase in the percentage of high conceptual level members in the group. Since the focus of the present study was upon group processes, the relation of the observer group functioning to group performance is uncertain. French (1951),

however, noted that equal participation of group members was positively related to group performance; a negative relationship existed between the degree of centralization and the level of performance. Similarly, Hutte (1965) has recently reported that groups performed more effectively when there was less leadership in regard to decision making.

Guetzkow's (1960) analysis of role differentiation and emergence of interlocking structures was based upon research involving communication nets. Effective performance in the various communication designs required role differentiation and, more importantly, the development of an interlocking organizational structure in the system, but, as Hutte has pointed out, the communication net research has been characterized by a topological channeling of information—a factor which was not present in the approach taken by Hutte or the present study and which may be one determinant of performance differences.

The underlying point here, then, is that different environments or task situations require different structural responses with respect to group organization. Granting that the task demands of the environment partially determine the structuring of the group, the present results suggest that the conceptual level of the group members is probably a limiting factor. The need for a low level of uncertainty, on the part of low conceptual level members in a complex environment, is reflected in the reduction of uncertainty with respect to differentiation and centralization of functional roles. The structural response may, then, be one source of personality-situation interaction effects.

The results of the present study confirmed the hypothesis (II) that groups, in which all of the members are of a high conceptual level, generate more interpersonal conflict than groups in which the members differ in conceptual level. Although there was a curvilinear relationship between generated conflict and the conceptual level dimension, the high levels of conflict represented by the end points of the curve appeared to be related to different processes. In the 100% high groups, the level of conflict was a function of the extensive evaluation of alternatives; although there was an *initial* increase in the

level of conflict when members attempted to support their proposals, extensive evaluation was conducive to conflict resolution. The conflict in the 25% high groups, on the other hand, was a function of the diversity generated; conflict resolution was attempted through persistent efforts on the part of the members to have their respective proposals accepted. The various alternatives were seldom assessed by other members or by the members responsible for them. Since the levels of conflict in the two different groups were similar but functions of different processes, differences in the extent to which the interpersonal conflict is utilized in decision making would not be unexpected (Hypothesis III).

The process of synthesizing or integrating perhaps relevant but discrepant information would seemingly have to be preceded by the process of evaluation. The increase in evaluation, which occurred with a decrease in group structuring, was paralleled by a linear increase in the group communication-channel complexity. A multichanneled communication network would be expected to facilitate reciprocal evaluation. Communication complexity consistently correlated highly with the rated utilization of conflicting information, the frequency of evaluation, and (negatively) with the S/E ratio. Although group structuring also correlated highly with these variables, the association was primarily attributable to the effects of communication complexity. It would seem that flexibility in performing different functions is a necessary but not a sufficient requirement for effective decision making; an openness to multiple sources of information is required on the part of each group member.

Another factor, which could be involved in the decrease of the S/E ratio with an increase in the proportion of highs, is the tendency of low conceptual level members to consider single rather than multiple alternatives, particularly under various stress conditions. This type of information processing provides a means of simplifying a situation to the point where it is possible for some action to be taken.

The hypothesis (IV) that the extent of novel information search increases with an increase in the proportion of highs in the

group was confirmed by the results. The increase in the proportion of information-search orders clearly paralleled the increase in the absolute number of search orders. Since the total number of decision orders did not significantly vary with group composition, a lower proportion of search orders for novel information would possibly indicate that the respective groups were more redundant in their search and retaliatory in their decisions; in other words, with a decrease in the percentage of members of a high conceptual level in the groups, there was a decrease in the use of long-range strategies. Informal observations would seem to be consistent with this interpretation.

REFERENCES

- BASS, B. M. *Leadership, psychology, and organizational behavior*. New York: Harper & Row, 1960.
- BENNIS, E. G., & SHEPARD, H. A. A theory of group development. *Human Relations*, 1956, 9, 415-437.
- COLLINS, B. E., & GUETZKOW, H. *A social psychology of group processes for decision-making*. New York: Wiley, 1964.
- EDWARDS, A. L. *Edwards Personal Preference Schedule*. New York: Psychological Corporation, 1959.
- FRENCH, J. R. P. Group productivity. In H. Guetzkow (Ed.), *Groups, leadership, and men*. Pittsburgh: Carnegie Institute of Technology Press, 1951. Pp. 44-54.
- GARNER, W. R. *Uncertainty and structure as psychological concepts*. New York: Wiley, 1962.
- GUETZKOW, H. Differentiation of roles in task-oriented groups. In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory*. Evanston, Ill.: Row, Peterson, 1960. Pp. 683-704.
- HEARN, G. The process of group development. Cassidy lecture, University of Toronto, 1955. (Mimeo) Cited in O. J. Harvey (Ed.), *Motivation and social interaction: Cognitive determinants*. New York: Ronald Press, 1963. P. 150.
- HUTTE, H. Decision-taking in a management game. *Human Relations*, 1965, 18, 5-20.
- KARLINS, M., SCHRODER, H. M., & STREUFERT, S. A controlled input program for a tactical war game. Washington, D. C.: American Documentation Institute, 1965, No. 8621.
- LANZETTA, J. T. Information acquisition in decision making. In O. J. Harvey (Ed.), *Motivation and social interaction: Cognitive determinants*. New York: Ronald Press, 1963. Pp. 239-265.
- MARTIN, E., & HILL, W. F. Toward a theory of group development: Six phases of therapy group development. *International Journal of Group Psychotherapy*, 1957, 7, No. 1. Cited in O. J. Harvey (Ed.), *Motivation and social interaction: Cognitive determinants*. New York: Ronald Press, 1963. P. 149.
- SCHRODER, H. M., DRIVER, M. J., & STREUFERT, S. *Human information processing*. New York: Holt, Rinehart & Winston, 1967, in press.
- SCHRODER, H. M., & HARVEY, O. J. Conceptual organization and group structure. In O. J. Harvey (Ed.), *Motivation and social interaction: Cognitive determinants*. New York: Ronald Press, 1963. Pp. 134-166.
- SIEBER, J. E., & LANZETTA, J. T. Conflict and conceptual structure as determinants of decision-making behavior. *Journal of Personality*, 1964, 32, 622-641.
- STAGER, F., & KENNEDY, J. L. Decision making and performance in heterogeneous and homogeneous groups. Technical Report No. 6, 1965, Princeton University, Contract Nonr 1858-(42), Office of Naval Research.
- STEINER, I. D. Group dynamics. *Annual Review of Psychology*, 1964, 15, 421-446.
- STREUFERT, S., CLARDY, M. A., DRIVER, M. J., KARLINS, M., SCHRODER, H. M., & SUEDELD, P. A tactical game for the analysis of complex decision making in individuals and groups. *Psychological Reports*, 1965, 17, 723-729.
- SUEDELD, P., & STREUFERT, S. Information search as a function of conceptual and environmental complexity. *Psychonomic Science*, 1966, 4, 351-352.
- TUCKMAN, B. W. Personality structure, group composition, and group functioning. *Sociometry*, 1964, 27, 469-487.
- TUCKMAN, B. W. The effects of group composition and task differences on group performance. Paper read at Eastern Psychological Association, New York, April 1966.
- WARE, C. T., JR. Individual and situational variables affecting human performance. *Human Factors*, 1964, 6, 673-674.

(Received March 28, 1966)

EFFECT OF MOTHER-CHILD SEPARATION AND BIRTH ORDER ON YOUNG CHILDREN'S RESPONSES TO TWO POTENTIALLY STRESSFUL EXPERIENCES¹

DAVID T. A. VERNON, JEANNE M. FOLEY, AND JEROME L. SCHULMAN

Children's Memorial Hospital, Chicago

2 experiments were carried out to investigate the effects of separation from mother and birth order on preschool children's responses to stress. The 2 potentially stressful situations were routine admission procedures and anesthesia induction. Separation from mother was not significantly related to children's responses to admission procedures, but did appear to make anesthesia induction more distressing, especially during the period just prior to induction and during the final moments of induction. The failure to find similar differences for the admission-procedures Ss may have been due to the fact that these procedures were found to be considerably less stressful than anesthesia induction. Neither birth order nor the combination of birth order and separation was related to responses to the potentially stressful situations. Also, in neither sample did the firstborn and later-born children differ in dependency nor did their mothers differ in the child-rearing practices assumed to contribute to dependency.

The variability of human responses to the threat of pain and bodily injury has been a matter of considerable interest to behavioral scientists. Much is now understood about why some persons remain calm in potentially stressful situations while others appear upset. The approaches and explanatory concepts, however, which have been used most successfully with adults (Chapman, 1962; Janis, 1958, pp. 3-39) do not appear to be applicable to very young children. For example, the psychological preparation and factual information about an impending stress which are demonstrably important for adults and older children are probably irrelevant for a 2- or 3-year-old.

The present experiments on children's responses to two hospital stress situations (admission procedures and anesthesia induction) direct attention to two variables which, in theory, are particularly salient for preschool children. These are (a) the child's separation from his mother during the stress experience,

and (b) the child's ordinal position in the family.

The following hypotheses about children's responses were tested:

1. Potentially stressful experiences are more distressing to children separated from their mothers than to children accompanied by their mothers.

2. Such experiences are more distressing to firstborn (including only) children than to later-born children.

3. The effect of separation is greater among firstborn children than among later-born children.

The first of these issues—the effect of mother-child separation during stress—while seemingly obvious, has not been thoroughly investigated. Arsenian (1943) found that children are less afraid of an unfamiliar playroom when in the presence of a familiar adult than when alone. However, her subjects, children of inmates in a women's reformatory, were probably quite special. The only additional data on mother-child contact during stress are from studies of relatively long separations (Fagin, 1964; Woodward, 1959, 1962). In addition, there is some indication that the mother's presence during stress may actually make the situation more upsetting. The findings of Campbell (1957) suggest that maternal anxiety about medical procedures is

¹Supported in large part by a grant from the Field Foundation, Incorporated.

The authors gratefully acknowledge the cooperation and assistance of the following persons: David Allan and the staff of the anesthesiology department of Children's Memorial Hospital, George S. Livingston, John Elsen, Paula Yonover, and Joan H. Schubert.

communicated to children, at least to infants under 6 months of age. However, this possibility was not supported by the findings of a similar study involving older children (Levy, 1960). The general trend of the above studies, along with the assumption that the mother is the major source of support and protection for the preschool child, led the investigators to conclude that her presence during stress should be comforting to the child.

With regard to the second and third hypotheses, studies of birth order demonstrate that firstborn persons (a) are more likely to be frightened by the threat of pain and physical harm, (b) have higher affiliative needs in such circumstances than do later-born persons, and (c) are more likely to benefit from affiliative activity during stress than are later-born persons (Sampson, 1965; Schachter, 1959). While these conclusions are based primarily on studies of young adults, careful consideration suggests that they should also be applicable to preschool children's responses to potentially stressful situations. The importance of early experience is suggested by Schachter's findings that firstborn persons and persons without siblings did not differ from each other in affiliative tendencies under stress conditions, but that both differed from later-born persons. This suggests that the social psychological processes leading to birth-order differences occur at an early age. Further, it seemed likely to the present investigators that the effects of later experience, which may attenuate birth-order differences, might be reduced by studying young children.

This line of reasoning was supported by the findings of a pilot study of children's responses to injections. The subjects were 28 hospitalized children between the ages of 1 and 6 years. Ratings of mood (made with the aid of a scale described later) revealed that the 6 firstborn children in this group were significantly more upset by the entrance of the medication nurse into the room than were the 22 later-born children. Observer bias was controlled by the fact that birth order was not known at the time of the ratings.

In addition to investigating the above hypotheses, the present experiments provided an opportunity to investigate several ideas concerning the origins of the hypothesized birth-

order differences. Schachter (1959) proposed that the reactions which are mediated by birth rank (i.e., anxiety and affiliation) are a result, in part, of variations in the parents' child-rearing practices. He suggested that parents' overprotection of the first child may contribute to the relatively high incidence of fear reactions among firstborn individuals. The idea that firstborn children are more dependent than later-born children received support from Haeblerle's (1958) data and data presented by Sears (1950).

While this line of reasoning is plausible, Sampson's (1965) review of evidence relevant to these relationships is not encouraging. Of the many ways in which parents' handling of firstborn and later-born children has been presumed to differ, only a few have been investigated, and, of these, the evidence frequently failed to support the posited relationship. The idea that the firstborn child is more dependent than the later born failed to obtain any clear-cut support.

The evaluation of the proposed relationships between birth order and reactions to threat, child-rearing practices, and dependency is further complicated by the fact that these variables have never been studied in a single sample. The present study, because of its focus on children's reactions to stress, provided an excellent opportunity for correcting this deficiency. Consequently, several measures of child-rearing practices and dependency were included in order to further investigate the following hypotheses:

4. Mothers of firstborn children more commonly use child-rearing practices presumed to contribute to dependency than do mothers of later-born children.

5. Firstborn children are generally more dependent than later-born children.

The selection of specific variables to be considered in the measures of child-rearing practices was based largely on frequently mentioned differences between parents of firstborn and later-born children (Sampson, 1965). On this basis, the assessment of mothers' activities with their children included measures of affection, protectiveness, encouragement of independence, permissiveness, and severity of control methods. Two additional variables, reward of dependency and with-

drawal of love, were included because of their apparent relevance to the development of dependency in preschool children (Sears, 1963; Sears, Maccoby, & Levin, 1957).

DESCRIPTION OF STUDIES

In testing the separation and birth-order hypotheses, it was assumed that the stressfulness of admission and anesthesia induction would be manifested in several areas of children's functioning: (a) mood during and immediately following the potentially stressful experience, (b) greater aggressiveness and less mature play immediately following the experience, and (c) changes in behavior suggesting increased upset following hospitalization. For practical reasons, it was not feasible to obtain data on these three aspects of behavior in both studies. Data on posthospital behavior were not collected in the study of admission procedures because it simply did not seem worthwhile—these procedures were of relatively minor stress in contrast to other aspects of these children's hospitalization. Instead, attention was given to play behavior immediately following admission. In the anesthesia study, the timing of surgery and the children's postoperative condition did not permit the play observations. Data on post-hospital behavior were added in their place.

EXPERIMENT I: ADMISSION PROCEDURES

The admission procedures investigated were those which were routinely performed in the first 15–30 minutes after the child's arrival on the ward. These were undressing and weighing and taking of rectal temperature and blood pressure.

Method

Subjects. The subjects were 32 children between the ages of 2 years and 5 years, 11 months. All were elective admissions, generally for surgery. Children with neurological disorders, acute illnesses, or traumatic injuries were purposely excluded, as were patients undergoing tonsillectomy.²

Procedure. The mothers of potential subjects were asked to participate at the time of their arrival at the hospital. Although participation was voluntary, only three mothers refused.

Prior to admission each child was classified with

² The frequency of various diagnoses was as follows: congenital cardiovascular disorders—15, hernia—6, other—11.

respect to age, sex, and birth order. Two age levels were used: young (2 years through 3 years, 11 months) and old (4 years through 5 years, 11 months). The two categories of birth order were firstborn (including only children) and later born. The first child of each composite type (e.g., young, male, later born) was randomly assigned to the accompanied or separated condition. The next child of the same composite type was assigned to the opposite treatment. The differences in conditions for the two groups were confined to the period of the admission routines themselves. For the accompanied children, these were carried out by a nurse with the mother at the child's bedside. In contrast, the mothers of the separated children were asked to leave the room for a few minutes while the child was prepared for his stay in the hospital. The nurse then carried out the admission routines alone. To help convince the children in the separated group that the separation was only temporary, the mothers left their personal belongings at the bedside.

Instruments and scoring. Data were obtained from behavioral observations and interviews with mothers about child-rearing practices.

Behavioral observations. Each subject was observed on his ward. In the manner of prior research on human responses to stress (e.g., Chapman, 1962; Janis, 1958) the period of observation was divided into four parts identified as the prethreat, threat, impact, and postimpact phases. The *prethreat phase* was the first 15 minutes after the subject's arrival on the ward. It was a period of free activity which each subject spent in a small playroom with his mother and an observer. Estimates of dependency, mood, quality of play, and aggression were based on observations of the child's behavior during this period. The *threat phase* began after the child returned to his room—at the time the mother left the child or would have left the child. The beginning of this brief phase was also marked by the entry of the admitting nurse into the subject's room. It ended with the start of the admission procedures. For the total sample it lasted an average of .6 minute. Ratings of mood were made during this and the following phase. The *impact phase* was the period during which the child underwent the admission procedures. It began when the nurse started to undress the child and ended when she removed the blood-pressure cuff. The average time required was 9.2 minutes. The *postimpact phase* was a second 15-minute period of free activity which was similar in all respects to that of the prethreat phase. It began approximately 5 minutes after the completion of the admission procedures.³

³ In order to facilitate the rating of mood, each of the four phases was subdivided into units as follows: (a) prethreat phase, each unit of action (as defined later for scoring quality of play) was rated separately, (b) threat phase, one unit, (c) impact phase, four units, corresponding to the four admission procedures, namely, undressing, weight-height, temperature, and blood pressure, (d) postimpact phase, same as prethreat.

One investigator (JMF) observed the subjects in the playroom and made a detailed record of mothers' and children's behavior. The observer took as little part as possible in the activities, but encouraged the mothers to interact with their children as they might at home. The observer was unaware of birth order and experimental condition (accompanied or separated during admission). To further minimize rater bias, the 64 records (Play 1 and Play 2) were arranged in random order and coded before being scored. Another investigator (DTAV) rated children's mood during the admission procedures.

Ratings of *mood* were based on the following 7-point scale:

1. Attentive and active in happy or contented way and/or interested in play or other constructive activity.
2. Attentive to surroundings, but not especially happy and/or unconstructive play and activity.
3. Passive, vacant expression, generally quiet, and no play. Dozing or sleeping.
4. Unhappy, worried, or anxious appearance without crying.
5. Marked unhappiness; whining, whimpering, moaning or soft crying; isolated indications of distress.
6. Moderate crying or intermittent but fairly constant sobbing; several noncrying reactions suggesting distress.
7. Scream full blast; intense and constant crying without paying attention to anything.

Four scores were calculated for each child. These were the mean mood ratings for the above four phases. To calculate these, the rating for each unit was weighted by the number of minutes covered by the unit, the products totaled, and then divided by the number of minutes in the phase. Previous work with this scale indicated a reasonable degree of agreement between independent observers. In a sample of 10 children the correlations (r) among the scores for the four phases ranged from .74 to .98.

Measures of dependency were made during the first (prethreat) play period; measures of quality of play and aggression were made during both the prethreat and postimpact phases. The *dependency scales* were based on criteria presented by Beller (1957), Gewirtz (1956), Haeberle (1958), Hartup (1960), Sears (1963), and Sears, Whiting, Nowlis, and Sears (1953). Of the various behaviors defined as dependency in these studies, five were selected for use in the play sessions, that is, seeking positive attention, seeking praise, seeking help and information, seeking comfort and reassurance, and seeking physical contact. The score for each measure was the number of minutes in which that particular activity occurred; that is, each aspect of dependency was rated 0 (absent) or 1 (present) for each of the 15 minutes of Play 1. Interjudge agreement for the five dependency scores was found to be reasonable for 16 children selected randomly from the present

TABLE 1
CORRELATIONS AMONG MEASURES OF DEPENDENCY

Seeking:	1	2	3	4
Positive attention	—	—	—	—
Praise	.14	—	—	—
Help, information	.54***	-.07	—	—
Reassurance	.25	-.21	.34*	—
Physical contact	-.08	-.14	.28	.56***

Note.— $N = 32$.

*.10 > p > .05.

*** $p \leq .01$.

sample. The correlations (r) between independent judges ranged from .74 to .94.

The correlations among the dependency scores (Table 1) were generally low and suggested that combining these measures into a single score was unwarranted.

Scores for *quality of play* were based on the rating procedures described by Foley (1962). The record of playroom behavior was first divided into natural units of action which were then rated on a 7-point scale. The lowest rating (1) was assigned to touching or holding a toy with minimal manipulation or examination; the highest (7) was assigned to highly elaborated and creative activities which were well sustained. The ratings for the units were then weighted for length and averaged. Prior work (Foley, 1962) with this scale in a 20-minute play session demonstrated that interscorer agreement was high and that the scores were responsive to variations in situational stress. (The same was true for aggression, described below.)

Aggressive acts (*aggression*) were defined as (a) those directed toward hurting another person, person surrogate, or animal, (b) destructive activities, and (c) assertive behavior, refusing to comply with rules and attempting to get own way (Foley, 1962). The score was the sum of the ratings when each aggressive action was rated 1 (or 2 if prolonged).

Interviews with mothers. These were usually obtained the day following admission, generally while the child was in the operating room or undergoing tests.⁴ The interview had two basic parts: the critical incidents interview and the parent activity questionnaire.

⁴ Whether the results of these interviews would be equivalent to data collected under completely neutral circumstances is not known. However, prior work with the interview for a different sample suggested that it is not significantly affected by variations in the degree of threat to which the child is exposed during hospitalization. Interviews made on the day of admission or the day following (when the children were either relatively ill or undergoing surgery) did not differ from interviews made 5 or more days after admission (when the children were well on the road to recovery) with respect to derived mean control methods scores ($t \leq 1.18$, $df = 40$, $p > .10$ in all cases).

The initial admission interview consisted of 14 questions probing admission history, child's behavior, family structure, and behavior. A child's record comprising a brief history on his behavior and admission history, "What is your best?" behavior, including admission to the hospital, were scored separately. Initial admission was scored for type and severity of admission incident. The "best behavior" category included compliance, initiative, and initiation of play or activity. The "worst behavior" category included failure, withdrawal, and physical punishment. The severity of post-admission incident was scored as low, moderate, or high (scored 1, 2, and 3).

The research project has several postadmission properties. First, admission may result in a child being severely injured, injured, or not injured. Any additional severe, withdrawal, or play and interest in hospitalization were also recorded. The score for withdrawal of play was the number of times the mother's response suggested that the child's behavior might endanger his attachment relationship with her. The number of threatened separations of the child's mother from his mother was scored as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11. The number of the five initial incidents which included separation during were scored as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11. The highest rating (11) indicated that the mother always thought the behavior presented a moderate severity that would keep the child away from her. The lowest rating (1) indicated that the mother tried to discourage or prevent the behavior by sending the child away or

removing him. The lowest rating (1) indicated that the mother always thought the behavior presented a low severity that would keep the child away from her. The highest rating (11) indicated that the mother always thought the behavior presented a high severity that would keep the child away from her.

The parent activity questionnaire concerned activities in three areas. The scale for affection included items about the parent's interactions with the child, for example, "play with child; take part in his games." The scale for encouragement included items about the parent's encouragement of the child, for example, "praise the child; encourage the child to do things he himself has done; encourage the child to play with you; encourage the child to play with other children." The scale for discipline included items related to encouraging the child to do things he himself has done, for example, "discipline the child; tell the child that he is doing badly; tell the child that he is doing well; tell the child that he is doing badly; tell the child that he is doing well."

There were 15 multiple-choice items for each scale. The mother checked the activity which best represented her usual activity with her child. The very often or almost always scored 4, fairly often scored 3, sometimes scored 2, rarely or never scored 1. The sum for each scale was the sum of the ratings for the 15 items comprising it. The total score for a mother was the sum of the scores for the three scales.

Results

In order to test the hypotheses about the children's responses to the admission procedures, data on mood, poststress quality of play, and poststress aggression were analyzed

TABLE 2
MOOD, QUALITY OF PLAY, AND AGGRESSION: MEANS AND RESULTS OF
ANALYSES OF VARIANCE AND COVARIANCE

	Mood				Quality of play		Aggression	
	Pre-threat	Threat	Post-impact	Post-impact*	Pre-threat	Post-impact*	Pre-threat	Post-impact*
<i>M's</i>								
Separation								
Separated (<i>N</i> = 16)	1.34	2.52	3.05	1.55	3.82	3.08	3.25	3.60
Accompanied (<i>N</i> = 16)	1.38	2.48	2.82	1.60	3.28	3.50	2.62	3.34
Birth order								
First born (<i>N</i> = 16)	1.45	2.59	2.86	1.60	3.51	3.27	2.56	2.12
Later born (<i>N</i> = 16)	1.26	2.41	3.01	1.56	3.60	3.30	3.31	4.82
<i>F</i> values								
Factorial analyses								
Separation (<i>df</i> = 1, 27)	—	.01	.17	.08	—	.95	—	.03
Birth order (<i>df</i> = 1, 27)	—	.15	.08	.04	—	.01	—	3.71*
Separation × Birth Order (<i>df</i> = 1, 27)	—	.91	.62	.48	—	.62	—	.10
One Way analyses								
Age (<i>df</i> = 1, 29)	—	1.89	2.81	3.63*	—	.01	—	1.13
Sex (<i>df</i> = 1, 29)	—	.62	.31	.04	—	.00	—	2.20
Prior hospitalization (<i>df</i> = 2, 28)	—	.38	.05	.03	—	.11	—	.45
Occupational status (<i>df</i> = 4, 26)	—	.58	.74	.91	—	.50	—	.71
Type patient (<i>df</i> = 2, 28)	—	10.09***	3.34**	.04	—	.82	—	.10

* Adjusted means computed by covariance.

* .10 > *p* > .05.

** .05 > *p* > .01.

*** *p* > .01.

TABLE 3
CHILD-REARING PRACTICES: MEANS FOR MOTHERS OF FIRSTBORN
AND LATER-BORN CHILDREN

	Activities					Parental Dependency					
	Affec- tion	Protect- ion	Love and affection	Control	Physical control	Physical dependency	Maternal dependency	Maternal dependency	Paternal dependency	Paternal dependency	Maternal dependency
Firstborn ($N = 16$)	47.75	48.75	45.81	142.31	3.81	48	2.00	2.16	55.44	2.38	1.94
Later-born ($N = 16$)	44.88	47.94	44.75	137.56	3.62	33	1.94	2.20	54.31	1.25	2.00
$F(2) = 1.30$	2.97*	14	30	1.29	.05	70	.43	.68	.08	3.95*	.03

* $10^{-5} < p < .05$.

by the covariance method as complete factorial experiments involving separation and birth order (2×2).⁵

Relationship of separation and birth order to children's responses. The mood scores for the threat, impact, and postimpact phases were analyzed separately with prethreat mood as the covariate. Quality of play and aggression during the play session following admission were also analyzed by the covariance method with data from the play session prior to admission used as covariates.

The analyses provided no support for the hypotheses. Neither separation nor birth order was significantly related (either singly or in combination) to the effects of the admission procedures by any measure. This may have been the result of the fact that the level of upset experienced by the subjects was generally low; in no phase did the mean mood score exceed 3.05 on the 7-point scale (see Table 2).

⁵ Initially it was intended to analyze the data as factorial experiments involving separation, birth order, age, and sex. Age and sex were included in order to increase the efficiency of the experiment. It was assumed that they would be related to the dependent variables. However, subsequent analyses indicated that this was not the case and that their inclusion in the factorial experiment served primarily to reduce the degrees of freedom in the denominator of the various F tests which were of special interest. For this reason age and sex were omitted in the factorial analyses. This simplified the resulting descriptions, but had no effect on the conclusions. In no case was a comparison statistically significant in the abbreviated (2×2) design while being nonsignificant in the full four-variable design. The procedures employed here are more conservative than "pooling" degrees of freedom.

The difference between firstborn and later-born children on aggression approached significance ($F = 3.71$, $.10 > p > .05$). However, contrary to the hypothesis, aggressive acts appeared to be somewhat less common among the firstborn than among the later born. Our initial thinking about children's responses to stress (based in part on the findings of Foley, 1962) led us to conclude that upset might be reflected in an increase in aggressive behavior following admission. However, data discovered later (summarized by Sampson, 1965) indicated that aggressive tendencies (and perhaps the tendency to express upset in aggressive acts) are generally lower among firstborn persons than among later borns. Such a tendency would, of course, confound the testing of the birth-order hypotheses of the present study.

Relationship of child-rearing practices and dependency to birth order. None of the F values was significant, but the differences between the birth-order groups on both affection and withdrawal of love were suggestive ($.10 > p > .05$). The finding that firstborn children received somewhat more affection than later-born children agrees with the popular notion that mothers are especially warm toward their first child. The greater use of withdrawal of love in controlling the behavior of firstborn children is consonant with the idea that that fear of loss of love, especially in an otherwise warm mother-child relationship, may lead to relatively high dependency.

The means and F values for the five measures of dependency for the birth-order groups (Table 4) did not support the hypothesis that

children are more dependent than later-born children.*

Relationship of other variables to responses. Age, sex, prior hospitalization, and occupational status (rated according to Coleman, 1959) were not significantly related to mood during the three phases or to quality of play and aggression after admission (Table 2).

The differences in mood for type of patient were statistically significant during both the threat and impact phases. The mean mood scores (adjusted) for the clinic, multiple-bed, and private patients during the threat phase were 2.39, 2.17, and 4.84, respectively. Corresponding means for the impact phase were 3.14, 2.50, and 4.61. The high level of upset exhibited by the children staying in private rooms appeared to be primarily a reflection of personality factors. Informal conversations with the mothers of these patients revealed that they had asked for private rooms so they could stay with their children during the night, feeling (in some cases on the basis of prior hospitalizations) that their children would be highly upset.

EXPERIMENT II: ANESTHESIA INDUCTION

The investigation of responses to anesthesia induction for tonsillectomy was of the same

*As one might expect, distributions of the dependency scores were quite skewed. In addition, in the analyses of variance, the cell means and the range of scores within cells tended to be positively related. Application of appropriate transformations (e.g., $y' = \sqrt{y + \sqrt{y + 1}}$) to several of the most skewed variables (both in this study and the study of anesthesia induction described later) corrected these conditions to a large extent, but had little effect on resulting F values and in no case affected the conclusions. For this reason only analyses of untransformed data are presented.

general design as the study of admission procedures.

Method

Subjects. The subjects were 32 children between the ages of 2 years and 5 years, 11 months. The children were admitted at 7:00 A.M., just prior to surgery. Thirty children were discharged in the late afternoon on the day of surgery; two remained in the hospital an extra day.

Procedure. The subjects were classified by age, sex, and birth order, and assignment to the separated and accompanied groups was made randomly. For the accompanied children, the induction was performed with the mother at the child's side. The mothers of the separated children were absent during induction. Both the accompanied and separated children experienced the same general conditions before and after surgery.

In describing the study to the parents, it was emphasized that participation was voluntary and that failure to participate would not affect the quality of the child's medical care. All parents who were asked participated.

The anesthesia used was a combination of halothane, oxygen, and nitrous oxide. All inductions were done in a special anesthesia room prior to entering the operating room for surgery.

Instruments and scoring. Data were obtained from behavioral observations, interviews with mothers, and a questionnaire about posthospital behavior.

Behavioral observations. The period before and during anesthesia induction was divided into four phases—prethreat, threat, and two impact phases (A and B). The *prethreat phase* was a 15-minute period of free activity following the subject's arrival in the hospital and prior to administration of the preoperative sedation. Each subject spent the period with his mother and the observer in a treatment room in the admitting area. The same play materials used in the admission study were provided. Estimates of dependency and mood were based on behavior during this phase. The *threat phase* was the interval between the time the mother left the child on the way to the surgical floor (or would have left the child) and the start of anesthesia induction. For the total sample the average length of this phase was 21.5 minutes. Only ratings of mood were made in

TABLE 4
DEPENDENCY: MEANS FOR FIRSTBORN AND LATER-BORN CHILDREN

	Dependency				
	Positive attention	Praise	Help, information	Reassurance	Physical contact
Firstborn ($N = 16$)	9.06	2.00	3.06	1.81	1.12
Late. born ($N = 16$)	6.31	1.56	1.56	1.88	.56
F ($df = 1, 30$)	2.37	.28	2.83	.01	.60

TABLE 5
MOOD AND POSTHOSPITAL BEHAVIOR: MEANS AND RESULTS OF
ANALYSES OF VARIANCE AND COVARIANCE

	Mood				Posthospital questionnaire—factor number						
	Prethreat	Threat	Impact A	Impact B	I	II	III	IV	V	VI	Total
<i>U</i> s											
Separation											
Separated	1.92	3.68	4.75	5.48	18.38	13.06	9.31	9.38	6.56	15.69	72.38
Accompanied	1.65	2.87	3.94	4.20	18.38	14.44	9.62	10.06	6.75	15.44	74.75
Birth order											
Firstborn	1.60	3.16	4.25	4.46	18.44	13.69	9.31	9.75	6.44	15.31	72.94
Later born	1.92	3.38	4.44	5.22	18.31	13.81	9.62	9.69	6.88	15.81	74.19
<i>F</i> values ^b											
Factorial analyses											
Separation $\cdot d^c$											
1, 27	—	3.13*	1.60	8.10***	.00	3.75*	.34	.66	.17	.20	.42
Birth order $\cdot d^c$											
1, 27	—	.23	.09	2.80	.10	.03	.34	.01	.96	.79	.12
Separation \times Birth Order $\cdot d^c$ 1, 27	—	1.18	.45	1.77	.92	.50	1.44	1.58	.02	.45	.07
One-Way analyses											
Age $\cdot d^c$ 1, 29	—	1.53	7.61***	.12	.00	1.48	2.95*	1.27	2.65	1.33	1.40
Sex $\cdot d^c$ 1, 29	—	.06	.01	1.08	.00	.26	.59	.44	.18	.05	.06
Mother anxiety $\cdot d^c$ 2, 12	—	1.02	.24	1.30	.59	.18	1.14	.27	.44	3.17	.12
Prior hospitalization $\cdot d^c$ 2, 28	—	2.66**	1.58	.44	.07	.32	.50	.38	.71	1.93	.83
Occupational status $\cdot d^c$ 4, 26	—	.40	1.90	.73	1.29	.84	.68	.60	.75	.43	.15
Anesthesiologist $\cdot d^c$ 6, 24	—	.51	2.32*	1.75	.92	.64	.70	1.50	2.53**	1.00	1.57

* Adjusted means compared by covariance.

^b Degrees of freedom listed are for covariance analysis (mood); degrees of freedom for analysis of variance (questionnaire) are decreased by 1 in the denominator.

* .10 > p > .05.

** .05 > p > .01.

*** p < .01.

this and later phases. *Impact Phase A* began with the start of induction when the mask was placed above the child's face and continued for 1 minute. *Impact Phase B* began with the end of Impact A and continued until a surgical level of anesthesia was achieved (marked by the injection of a muscle relaxant).⁷ The average length of the phase was 1.9 minutes.

The mood scores for the four phases and the ratings of dependency for the prethreat play period were based on the scales and scoring techniques described earlier.

Interviews with mothers. The methods and scoring were identical to those of the admission study.

⁷ The phases were subdivided into the following units: (a) prethreat phase, units of action as described for the admission study; (b) threat phase, three units—the first was the period from the time the mother left (or would have left) until the time the child entered the anesthesia room; the second was from the entry until the stethoscope was taped to the child's chest; the third was from that time until the mask was positioned above the child's face; (c) Impact Phases A and B, one unit each.

Questionnaire on posthospital behavior. The questionnaire used to assess the effects of separation consisted of 28 items (e.g., "Does your child seem to be afraid of leaving the house with you?; Does your child follow you everywhere around the house?; Is your child afraid of the dark?"). It was sent to the parent 6 days after discharge. For each item the respondent compared the child's typical behavior in the first week after hospitalization with his typical behavior before hospitalization. Five response alternatives were provided: (a) much less than before—scored 1, (b) less than before—scored 2, (c) same as before—scored 3, (d) more than before—scored 4, and (e) much more than before—scored 5.

On the basis of an earlier factor analysis of the questionnaire (Vernon, Foley, & Schulman, 1966), six types of responses to hospitalization (and a total score) were distinguished: I—general anxiety and regression, II—separation anxiety, III—sleep anxiety, IV—eating disturbance, V—aggression toward authority, and VI—apathy and withdrawal.

Prior work on validity and test-retest reliability was encouraging (Casell, 1965; Vernon et al., 1966).

TABLE 6
CHILD-REARING PRACTICES: MEANS FOR MOTHERS OF FIRSTBORN
AND LATER-BORN CHILDREN

	Anxious				Control practices						
	Affect them	Protect them	Inter- fere	Isolate	Reassure (encourage)	Proper discipline	Mean dependency score (unweighted)	Mean dependency score (weighted)	Total severity	With- drawal of love	Reward dependency
Firstborn (<i>N</i> = 16)	47.25	46.31	45.81	188.31	3.44	33	2.04	2.14	56.94	2.00	1.75
Later born (<i>N</i> = 16)	45.25	47.31	45.44	189.06	3.25	39	2.09	2.22	59.69	2.06	1.69
<i>F</i> (<i>df</i> = 1, 30)	1.34	.24	.03	.05	.06	1.82	.28	1.69	.44	.02	.04

Results

The analyses in this experiment paralleled those of the study of admission procedures. The mood scores for the four phases were analyzed separately with prethreat mood as the covariate. The data on posthospital behavior, dependency, and child-rearing practices were analyzed by analyses of variance. The posthospital behavior questionnaires were scored for the six types of responses to hospitalization and a total score.*

Relationship of separation and birth order to children's responses. Two of the three analyses of covariance for mood provided some support for the hypothesis that the experience is more distressing for children separated from their mothers than for children accompanied by their mothers. The difference was statistically significant for Impact Phase B and approached significance for the threat phase.

Similar differences between the accompanied and separated groups were not obtained for

the period following hospitalization. The one difference which approached significance indicated that children who had been accompanied by their mothers scored somewhat higher in separation anxiety (Factor II) than those who had been separated. This finding was contrary to expectations since increases in clinging and attention getting (behaviors represented by this factor) have commonly been interpreted as a result of separation during hospitalization (Schaffer & Callender, 1959).

The hypotheses about birth order received no support in any of the analyses.

Relationship of child-rearing practices and dependency to birth order. The means and analyses of variance (Table 6) for the child-rearing-practices scores indicated that mothers' reports of their activity with their children and their control methods did not vary on the basis of birth order. The possibility that mothers of firstborn and later-born children differed with respect to affection and withdrawal of love, as suggested by the findings for the admission sample, was not confirmed.

The mean scores for dependency and the results of the one-way analyses of variance

TABLE 7
DEPENDENCY: MEANS FOR FIRSTBORN AND LATER-BORN CHILDREN

	Dependency				
	Positive attention	Praise	Help, information	Reassurance	Physical contact
Firstborn (<i>N</i> = 16)	9.06	1.25	1.75	2.00	2.56
Later born (<i>N</i> = 16)	7.31	1.38	1.56	1.56	3.12
<i>F</i> (<i>df</i> = 1, 30)	.88	.02	.06	.43	.10

* Because the parents of two children did not return usable questionnaires the method of unweighted means was used (Winer, 1962, pp. 241-244).

are reported in Table 7. As with the admission phase, these analyses did not support the hypothesis that first-born children are more dependent than later-born children.

Relationship of other variables to responses. Several of the variables considered were completely unrelated to the children's responses by any measure (Table 5); these were the sex of the child, the level of the mother's anxiety during induction (rated high, moderate, and low for mothers who accompanied their children), and occupational status.⁹

Three variables were significantly related to the children's responses at one time or another; these were prior hospitalization, the particular anesthesiologist or anesthetist handling the induction, and age (Table 5). During the threat phase those children who had been in the hospital before appeared to be slightly more upset than those who had not. The mean mood scores (adjusted) for children with no prior hospitalization, with 10 days or less prior hospitalization, and with over 10 days prior hospitalization were 2.99, 3.64, and 5.50, respectively. These differences were felt to reflect variations in the children's ability to recognize that the situation was threatening—an ability which would logically be related to the extent of prior experience in such settings.

Differences among the seven anesthetists who participated in the experiment appeared to be important on two measures: (a) mood during the first minute of induction (Impact Phase A), and (b) aggression toward authority following hospitalization (Factor V). The differences in mood between the children handled by each anesthetist (Table 8) agreed with the impressions of the investigator who observed the inductions. It seemed that these differences were due primarily to differences in the extent to which the anesthetist could (or tried to) establish rapport with the children in the few minutes prior to induction.

The difference in mood between young and old children was significant during the first minute of induction (Impact Phase A). At

⁹ Type of patient, a classification which was found to encompass some significant differences in the study of admission procedures, was not analyzed in this experiment because none of the patients was in a private room.

TABLE 8
MOOD (IMPACT PHASE A) AND POSTHOSPITAL BEHAVIOR (AGGRESSION): MEANS OF CHILDREN HANDLED BY DIFFERENT ANESTHETISTS

Anesthetist	N	Impact Phase A		Posthospitalization questionnaire	
		Mood		Aggression	
		M	Rank	M	Rank
A	7	4.95	3	6.86	2
B	3	5.75	2	8.33	1
C	4	4.11	4	6.75	3
D	3	2.53	7	5.50	6.5
E	9	3.63	6	6.56	4
F	3	3.85	5	6.00	5
G	3	6.30	1	5.50	6.5

this time the 2- and 3-year-olds appeared to be considerably more upset than the 4- and 5-year-olds; the mean mood scores (adjusted) being 5.26 and 3.43, respectively. The differences in mood during the other phases were similar in direction, but were less pronounced. Age differences approached significance on sleep anxiety (Factor III): the young children had a mean of 9.9 and the older children a mean of 9.0.

DISCUSSION

Separation

The hypothesis that separation from mother makes potentially stressful situations more distressing received some, but not unequivocal, support. This came primarily from data on behavior during the latter part of anesthesia induction and, to a lesser extent, from data on behavior immediately prior to induction.

There was no indication, however, that the mother's presence affected posthospital behavior. Although several of the children appeared to experience considerable difficulty in the week following hospitalization, it was not demonstrably related to separation during induction. The effects of separation may well have been outweighed by other stressful experiences during hospitalization.

The fact that the hypothesis about the effects of separation received support from the study of anesthesia induction but not from the study of admission procedures may have been due to differences in the stressfulness of the two situations. The level of upset shown

during the admission procedures was relatively low during all phases in comparison with anesthesia induction (see Figure 1). As a result, the average rating of mood for the entire group during the impact phase was approximately 3—a level which provides no clear indication of upset or fear. Another indication that the admission procedures were not generally stressful lies in the comparison of the means for quality of play and aggression before and after admission. The very small changes found here, in contrast to the significant differences found by Foley (1962), suggest that the situation was not inherently distressing.

It is interesting to note that the effect of separation during induction was greatest just prior to sleep—during the so-called "excitement phase." This may be due to the fact that this was the most stressful period of induction because of the likelihood of frightening physical sensations or because self-control was relatively low with a corresponding increase in emotional expression. Throughout these experiments, the differences between the separated and accompanied children appeared to vary with the overall level of stress, the

greater differences being associated with the more stressful phases. This is in line with Schachter's original data which indicated that the need of affiliation is high under conditions of relatively high stress. In addition, the present results extend this finding from the psychological laboratory to a real-life stress situation and suggest that it is applicable to children as well as young adults.

Birth Order

The hypotheses concerning birth order received no support. In neither experiment was there any clear indication that firstborn and later-born children differed in the extent to which admission or anesthesia induction was distressing; this was true for both the separated and the accompanied children.¹⁰

The fact that birth order was not related to responses was surprising since, as discussed earlier, Schachter's theory should be applicable to children's reactions to medical procedures. Such procedures threaten pain and physical harm and are the type of threat for which past evidence has been most conclusive.

In addition, the child-rearing practices reported for the firstborn and later-born children did not differ consistently in ways presumed to promote dependency (or in any other way) nor did these two groups differ significantly in dependency. For the present experiments, the data on birth order present a consistent and logical picture: Mothers of firstborn and later-born children did not differ in their use of child-rearing practices presumed to contribute to dependency; therefore, firstborn and later-born children were not different in dependency and, consequently, did not respond differently to separation from their mothers during a stressful experience.

¹⁰ This statement may be carried even further on the basis of results from the full four-variable factorial design noted earlier. In this analysis separation, birth order, age, and sex were considered together. There was no indication that the predicted effect of birth order alone or the predicted interaction between birth order and separation was to be found within either the age or sex subgroups. This was true for both samples. In only two instances (both in the admission sample) were there significant higher order interactions involving birth order. In neither case was there a subgroup pattern of the predicted type which could be replicated in the anesthesia-induction sample.

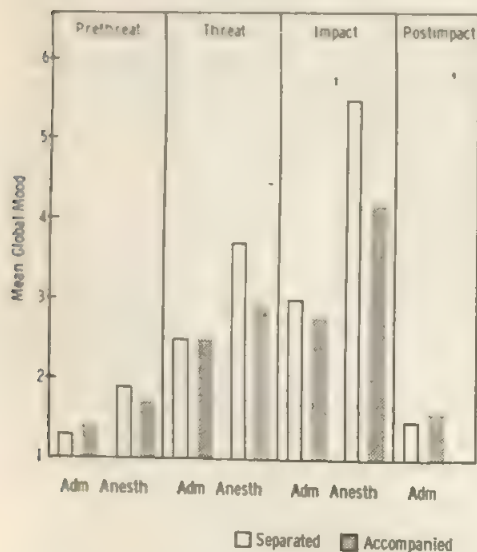


FIG. 1. Mean mood scores (unadjusted) for the admission-procedures and anesthesia-induction samples during different phases. (For the anesthesia sample only Impact B is shown. No postimpact data on mood were collected on this sample.)

It is possible, however, that the particular samples which were studied may have affected the birth-order findings. The sample for the study of admission procedures contained many children who were chronically ill. This factor may have affected the mothers' child-rearing practices in ways which attenuated the predicted birth-order differences. However, this was certainly not the case for the study of anesthesia induction. Although these children seemed quite normal and reasonably healthy, the differences between the firstborn and later-born children were smaller for this sample.

A final possibility is that the children in these experiments were too young. This might be the case if the hypothesized differences for birth order were the result of relationships with siblings and peers. For example, the concept of the sibling as a model (Sampson, 1965) suggests that firstborn children have only their more distant and powerful parents for models. They lack a peer in their homes with whom they can identify. This initial lack of a peer model may lead the firstborn to seek affiliation with others in order to find a model who will be of assistance in understanding himself. If this factor is important in the development of affiliative needs, the present subjects may simply have been too young to have established these patterns of behavior. Further, if the need for self-understanding (as satisfied through the social comparison process) is a determinant of birth-order differences, it is possible that the affiliative activities of firstborn persons might be directed more towards peers than parents. In this event, the presence or absence of mothers might be inappropriate for assessing differences related to birth order.

General Comments

The following comprehensive picture of children's responses to stress is suggested by the present data: It is a complex process in which the determinants of individual differences are not the same for all phases of the stress experience, but, rather, vary with the level of stress and the presence of supports or constraints on expression.

When the level of potential stress is low, as during the threat phase of admission, dif-

ferences in response are primarily a matter of personality. Only those children who are unusually sensitive to the threat of the situation, for example, the children in private rooms, are upset. The great majority are not upset, and the supports available in the setting, particularly the presence of the mother, are not generally needed and, consequently, have no measurable effect.

As the level of potential stress increases, or as the situation becomes one which many persons recognize as threatening on the basis of prior experience, the potential value of various psychological supports increases. Thus, during the threat phase of anesthesia induction the presence of the mother appeared particularly important. If, at this level, or a slightly more stressful level, other supports, such as the friendliness of the anesthesiologist, are available, they may influence the child's response.

As the level of potential stress becomes still higher and the ability to control emotional expression is reduced, as in the latter part of anesthesia induction, only those supports which are most basic or are most highly valued (e.g., the presence of the mother) are capable of influencing the response or making the situation less stressful. In the present study a difference between the separated and accompanied children was found in the threat phase of anesthesia induction because the mother was the only support available. In contrast, the difference between these groups during the latter part of induction was significant because the presence of the mother was the only support which was effective.

REFERENCES

- ARSENIAN, J. M. Young children in an insecure situation. *Journal of Abnormal and Social Psychology*, 1943, **38**, 225-249.
- BARGER, P. Parental dominance and conflict—relationship to personal history variables and child control techniques. Unpublished doctoral dissertation, Northwestern University, 1963.
- BELLER, E. K. Dependency and autonomous achievement striving related to orality and anality in early childhood. *Child Development*, 1957, **28**, 287-315.
- CAMPBELL, E. H. Effects of mothers' anxiety on infants' behavior. Unpublished doctoral dissertation, Yale University, 1957.
- CASSELL, S. Effect of brief puppet therapy upon the emotional responses of children undergoing cardiac

- catheterization. *Journal of Consulting Psychology*, 1965, 29, 1-8.
- CHAPMAN, D. W. A brief introduction to contemporary disaster research. In G. W. Baker & D. W. Chapman (Eds.), *Man and society in disaster*. New York: Basic Books, 1962. Pp. 3-27.
- COLEMAN, R. P. Social class in Kansas City. Unpublished doctoral dissertation, University of Chicago, 1959.
- FAGIN, C. The case for rooming in when young children are hospitalized. *Nursing Science*, 1964, 2, 324-333.
- FOLEY, J. M. Children's responses to frustration: Relationship with child rearing practices and infant reaction patterns. Unpublished doctoral dissertation, Northwestern University, 1962.
- GEWIRTZ, J. L. Dependent and aggressive interaction in young children. Unpublished doctoral dissertation, State University of Iowa, 1956.
- HAPPERD, A. Interactions of sex, birth order, and dependency with behavior problems and symptoms in emotionally disturbed preschool children. Paper read at Eastern Psychological Association, Philadelphia, April 1958.
- HARTUP, W. W., & KELLER, G. D. Nurturance in preschool children and its relation to dependency. *Child Development*, 1960, 31, 681-689.
- JANIS, I. *Psychological stress*. New York: Wiley, 1958.
- LEVY, D. M. The infant's earliest memory of inoculation: A contribution to public health procedures. *Journal of Genetic Psychology*, 1960, 96, 3-46.
- SAMPSON, E. F. The study of ordinal position: Antecedents and outcomes. In B. A. Maher (Ed.), *Progress in experimental personality research*, Vol. 2. New York: Academic Press, 1965. Pp. 17-33.
- SCHACHTER, S. *The psychology of affiliation*. Stanford: Stanford University Press, 1959.
- SCHAFER, H. R., & CALLENDER, W. W. Psychological effects of hospitalization in infancy. *Pediatrics*, 1959, 24, 528-539.
- SEARS, R. R. Ordinal position in the family as a psychological variable. *American Sociological Review*, 1950, 15, 397-401.
- SEARS, R. R. Dependency motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation*, 1963. Lincoln: University of Nebraska Press, 1963. Pp. 25-64.
- SEARS, R. R., MACCOBY, E. R., & LEVIN, H. *Pattern of child rearing*. Evanston, Ill.: Row, Peterson, 1957.
- SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, P. S. Some child rearing antecedents of aggression and dependency in children. *Genetic Psychology Monographs*, 1953, 47, 125-255.
- VERNON, D. T. A., FOLEY, J. M., & SCHULMAN, J. L. Changes in children's behavior following hospitalization: Some dimensions of response and their correlates. *American Journal of Diseases of Children*, 1966, 111, 581-593.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.
- WOODWARD, J. Emotional disturbances of burned children. *British Medical Journal*, 1959, 1(5128), 1009-1013.
- WOODWARD, J. Parental visiting of children with burns. *British Medical Journal*, 1962, 2(5320), 1656-1657.

(Received April 5, 1966)

VERBAL CONDITIONING AND CHANGE ON PERSONALITY MEASURES¹

MORTON G. HARMATZ

University of Massachusetts

Recent studies have attempted to demonstrate that changes in verbal behavior produced by verbal conditioning techniques could lead to changes on postconditioning personality tests. The method has been to increase the frequency of a particular response class (usually positive and negative self-references) through positive reinforcement (ummm-hmm). The results have been equivocal. Increasing the frequency may be strengthening the existing verbal repertoire rather than changing that repertoire. This may have the effect of reducing the likelihood of demonstrating changes on postconditioning personality tests. A conditioning procedure was presented and evaluated which manipulated the degree of endorsement S indicated to positive or negative self-references. Resulting changes on a semantic differential, a Q sort, and personality inventories were evaluated. The results indicated that the conditioning procedure was effective in manipulating verbal behavior, and changes occurred on many but not all, of the postconditioning personality tests.

The literature on verbal conditioning has demonstrated the efficacy of conditioning techniques in the manipulation of verbal behavior. In addition, this research has delineated some of the relevant parameters involved in verbal conditioning (Krasner, 1958; Salzinger, 1959). One question which has not received much attention, but which is of interest to many experimenters in this area, concerns changes in personality as a result of changes in verbal behavior produced by conditioning procedures. This question is a direct concern of those involved in psychotherapeutic investigations where verbal interaction between a patient and therapist is supposed to lead to personality change in the patient.

The research relating verbal conditioning to changes in personality includes three studies, all of which employed paper-and-pencil personality tests as their measure of personality. Nuthmann (1957) employed a true-false personality test of self-acceptance especially constructed for her experiment. Using subjects who had received low scores on this test, she had each subject retake the test in an individual session where he re-

sponded by throwing a switch true or false. Each answer reflecting acceptance of self was reinforced by "good" in one group, and a light blink in a second group. She found significant changes in a more self-accepting direction for those subjects reinforced with good.

Koenig (1962) was interested in changing scores on the Test Anxiety scale (TAS) by selective reinforcement of the person's statements of positive or negative self-statements about academic situations. The conditioning session was a free-verbalization one restricted to the subject's feelings about himself as a student. In one group Koenig reinforced positive academic self-statements, in another, negative academic self-statements. He found that the group reinforced for negative self-statements decreased significantly in TAS scores. No change was found for those reinforced for positive self-statements.

Rogers (1960) attacked this problem quite directly by employing a quasitherapy situation. He conditioned positive and negative self-references in six 10-minute quasitherapy interview situations in which subjects talked freely about themselves. This was preceded and followed by a Q-sort test. Subjects were divided into three groups, one reinforced for positive self-references, one reinforced for negative self-references, and a nonreinforced control group. The results showed clear-cut

¹ This paper is based on a dissertation submitted to the University of Washington in partial fulfillment of the requirements for the PhD. The writer is indebted to his sponsor, Irwin G. Sarason, and to Allen Edwards, Ezra Stotland, and Samuel Himmelfarb.

conditioning of negative self-references. Positive self-references extinguished under the control condition, and this extinction was arrested by reinforcement. No change, however, occurred on the personality tests as a function of conditioning.

Koenig and Rogers, as most other investigators of verbal conditioning, employed the frequency of the desired response class as the dimension of the response to be increased through conditioning. This dimension is appropriate in many of the verbal conditioning studies. However, in any study attempting to effect a change in a personality measure, this may not be the appropriate dimension of the response to employ. The assumption of such a manipulation is that the verbal repertoire has been changed so that the person is responding *differently* than he had prior to the conditioning procedure and that these changes will now generalize to the personality measure. In frequency manipulations, such as Rogers', the subject emitted those positive and negative self-statements which were already high in his verbal repertoire. Rogers reinforced these responses, thus strengthening the existing repertoire. It would appear that the subject would continue to make the same positive or negative self-references which were high in the repertoire before conditioning and which would be included on a personality test which assessed these references in some way. Changes should *not* be expected after conditioning with this procedure.

PROBLEM

It is the intention of this study to demonstrate and evaluate a conditioning procedure which avoids the problems of conditioning the *frequency* of the response class. In addition, this procedure will be examined with regard to its effectiveness in changing responses to personality tests after conditioning.

The major problem is how to initiate a new response and make it part of the verbal repertoire rather than strengthening existing responses. If the subject is given a self-reference on a true-false test, he could accept or reject the item according to his existing repertoire. If the items, however, were presented on a 9-point scale ranging from true

to false, most items would fall at some point other than completely true or completely false. In addition, the point where the item is placed by the subject will be somewhat indefinite from one presentation to another. If the subject answers the same class of items over and over, and his answers are sometimes more toward the true end or false end of the scale, it should be possible to reinforce, systematically, changes in a given direction for a particular class of responses. In this manner the degree of endorsement necessary for being reinforced, with regard to a specific class of items, could be successively shifted toward the true or false end of the scale. That is, the subject could be "shaped" into more extreme acceptance or rejection of the desired item class.

This procedure was employed with positive self-references and negative self-references as the item classes. Since positive self-references and negative self-references are opposites, a further possibility is to reinforce a group for accepting positive self-references as true of the self, while simultaneously reinforcing them to more strongly reject negative self-references as being true of the self, and vice versa.

The assessment of the generalization to personality tests was measured through three personality measures administered before and after conditioning. These included semantic differential rating scales, a *Q* sort, and a battery of true-false personality tests.

METHOD AND PROCEDURE

Subjects

The subjects were 50 (25 males, 25 females) students from the introductory psychology course at the University of Washington. They were all between 17 and 22 years of age. To avoid subjects who dealt with self-references in an extreme manner, the Edwards (1957) Social Desirability (*SD*) scale was administered, and subjects were selected from the middle 20% of the *SD*-scale distribution.

Preconditioning Tests

The introductory psychology class was administered a battery of personality tests including the *SD* scale. The battery (Sarason, 1958) included the following scales: TAS, General Anxiety scale (GA), Hostility scale (H), Lack of Protection scale (LP), and Defensiveness scale (D). This administration

occurred 1-2 months prior to the first conditioning session

The semantic differential and the Q sort were administered to the subject individually immediately before the first conditioning session. The semantic differential rating scales (Osgood, Suci, & Tannenbaum, 1957) required the subject to rate a concept along a given number of polar scales. The concepts which were rated included: Myself, Me As I Would Like to Be (the ideal self), Me As Others See Me (social self), Death, The Good Me, The Bad Me, The Experimenter. These concepts were rated along the following polar scales: large-small, active-passive, clean-dirty, strong-weak, slow-fast, light-heavy, fair-unfair, hot-cold, good-bad. These scales were chosen for their high loading on one factor and negligible loading on the other two factors. The factors used were the evaluative, potency, and activity factors.

The Q sort was adapted from one employed by Butler and Haigh (1954). It was composed of 50 statements (25 positive self-references and 25 negative self-references) each typed on a separate card. The subjects had to sort the items into nine categories from "least like me" to "most like me" according to a forced normal distribution.

Conditioning Sessions

Each subject was given a stack of 3×5 cards, each of which contained a self-reference (positive, negative, or neutral) and a scale with numbers ranging from 1 to 9 as follows:

Very Untrue										Very True
	1	2	3	4	5	6	7	8	9	

The subject was presented with a sample card which read: "I like movies" and these instructions:

We will now go through some cards like the one you are holding. Each card will have a statement and a scale. What you are to do is to tell me how true or untrue a particular statement is of you by reading the statement aloud and then calling out a number from one through nine. As you can see the one end means the item is very untrue of you, while the nine end means it is very true for you. Do you have any questions?

The subject then went through the stack of cards reading the statement aloud and calling out a number.

There were 120 items consisting of 52 positive self-references, 52 negative self-references, and 16 neutral self-references. The subject went through these items twice in the first session and a third time in the second session. The order was changed for each presentation of the items.

Treatment Groups

The first variable in this study compares subjects reinforced for accepting positive self-references as true of the self with those who were reinforced for accepting negative self-references as being true

of the self. Acceptance is defined as placing an item toward the true end of the scale and rejection, as placing the item toward the false end of the scale. The designations for these groups are: +PSR (for subjects reinforced for accepting positive self-references), and +NSR (for subjects reinforced for accepting negative self-references).

The second variable examines the direction of reinforcement. It was noted above that since positive and negative self-references are opposites, and since the task employed contained both types of items, it was possible to reinforce acceptance of one type of self-reference and rejection of the other type. Due to an absence of previous research, however, it was impossible to predict how such a procedure might effect conditioning. For this reason it was thought best to compare subjects under the simpler conditioning (toward the true end of the scale only) with subjects who were reinforced in both directions, that is, toward the true end of the scale on one type of self-reference, and toward the false end of the scale on the other. These groups are designated: T (for those subjects reinforced in the true direction only), and T & F (for those subjects reinforced toward the true and false ends of the scale).

The first and second variables each have two levels. The four possible combinations of these variables and the procedures for each of them during the conditioning session are as follows:

+PSRT. Subjects were reinforced for placing positive self-references toward the true end of the scale. This was accomplished in the following manner: every time a response to a positive self-reference was given at the 5 or higher level of the scale, the experimenter said, "um-hmm," until 15 reinforcements of this nature were given. At this point the subject had to give a response which was at the 6 or higher level of endorsement before a reinforcement was given, and this was continued until 15 reinforcements were given. The level then was moved up to 7 and higher responses, for an additional 15 reinforcements, and finally to 8 and higher. In this manner, the endorsement of a particular class of items, in this case positive self-references, was "shaped" to stronger and stronger "true" endorsements.

+NSRT. Subjects were reinforced for placing negative self-references toward the true end of the scale. This was accomplished exactly as in the +PSRT group above, but in this case the reinforced class of items was negative self-references.

+PSRT, +NSRF. Subjects received reinforcement for placing positive self-references toward the true end of the scale and for placing negative self-references toward the false end of the scale. This group was reinforced for positive self-references exactly as the first group (+PSRT) above, but in addition all responses to negative self-reference items of 5 and lower were reinforced with "um-hmm" until 15 reinforcements were given, at which time the level was moved to 4 and lower responses for another 15 reinforcements, then to 3 and lower,

and finally to 2 and lower. This group, then, while being reinforced for stronger and stronger acceptance of positive self-references, was simultaneously being reinforced for stronger and stronger rejection of negative self-reference items as being true of self.

+NSRT, +PSRF. Subjects were reinforced for placing negative self-references toward the true end of the scale and positive self-references toward the false end of the scale. This was done exactly as in the +PSRT, +NSRF group above, but in this case the negative self-references were moved toward the true end and positive self-references toward the false end of the scale.

Finally, a fifth group, the O group, was a control group which went through the same procedures as the above four groups, but received no reinforcement. Subjects were randomly assigned to each of these five groups with males and females equally represented in each group.

To review the procedure of Day 1, the subject took the Q-sort test, filled out a semantic differential booklet, and went through a conditioning procedure (or nonconditioned control procedure).

When the subject returned on Day 2, he was started out on a set of cards which again were the same as the original set, but in a third random order. The conditioning procedure was exactly the same for each subject as on Day 1 except reinforcement started at the 6 level for reinforcement in the true direction and at the 4 level for reinforcement in the false direction. It continued from this point in the same manner as on Day 1.

After conditioning, the subject was once again given the Q sort and the semantic differential. When these were completed, the subject was handed a mimeographed scale and IBM scoring sheet and told to answer the items. The experimenter left the room while the subject worked on the scale. The scale contained the five personality scales which had been used at the beginning of the quarter, that is, TAS, GA, H, LP, and D. The experimenter returned and made sure the subject had answered all items. Some of the subjects were also given an awareness interview adopted from Rogers (1960). The subject was then dismissed with the experimenter's thanks

and with a request that the subject not mention the procedure to his classmates.

RESULTS

The scores for the analyses which follow were derived by summing the numerical response the subject had given (1-9) separately for responses to positive self-reference items, negative self-reference items, and neutral items. These sums were then multiplied by the score categories yielding a sum of the ratings for each of these item classes.

The analyses were repeated-measurement analyses of variance examining the variables of conditioning, direction, and sex of subject. The conditioning variable compared those subjects conditioned for positive self-references (+PSR) to those conditioned for negative self-references (+NSR) and thus has two levels. The direction variable compared those conditioned toward the true end of the scale (T) with those conditioned toward the true end for one response class and toward the false end on the other response class (T & F), as was previously described and, therefore, also has two levels. The sex variable, of course, has two levels: male and female.

Analysis of Ratings to Positive Self-Reference Items

The first analysis was a $2 \times 2 \times 2 \times 3$ repeated-measurement analysis of variance for the ratings to the positive self-reference items. The major finding here was for the variable of conditioning. The F ratio for conditioning of 7.29 was indeed significant ($df = 1, 32$, $p < .025$). This difference is accounted for by the difference between those subjects reinforced for accepting positive self-references (+PSR), and those subjects reinforced for accepting negative self-references (+NSR). That is, those subjects reinforced for positive self-references had significantly higher scores on positive self-references ($\bar{X} = 932.90$) than did those subjects reinforced for negative self-references ($\bar{X} = 829.90$). It is clear from this finding that conditioning did effect the verbal behavior of the subjects in the predicted direction.

A closer look at the conditioning groups across trials is given in Figure 1. Those sub-

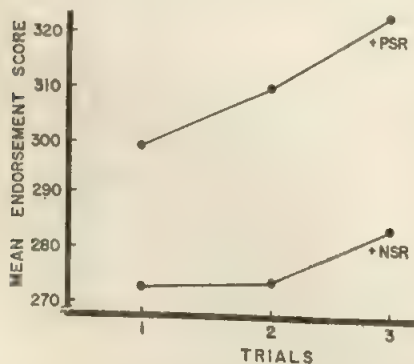


FIG. 1. Mean endorsement scores to positive self-reference items for +PSR and +NSR groups.

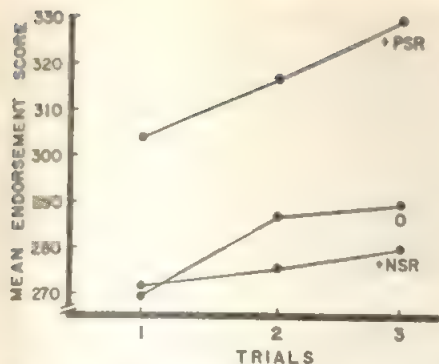


FIG. 2. Mean endorsement scores to positive self-reference items for +PSRT, +NSRT, and control groups.

jects reinforced for positive self-references were significantly,² different on all three trials from those subjects reinforced for negative self-references. In addition, the Trial II +PSRT group was significantly higher than the Trial I +PSRT group, and the Trial III +PSRT group was higher than the Trial II +PSRT group. Thus each trial represented a significant rise over the preceding trial. The Trial I +NSRT and the Trial II +NSRT groups did not differ. The Trial III +NSRT group was, however, significantly different from both Trial I and Trial II +NSRT groups.

In order to compare these groups to the nonreinforced control group (O) it was necessary to separately show the T groups and the T&F groups since there can be no "non-reinforced" control for direction. Figure 2 shows the curves and the scores for the T groups. The +PSRT group was significantly higher than the O control on every trial. However, the +NSRT group was not significantly lower than the control group.

This same analysis performed on the T&F groups showed a similar picture. The +PSRT group was significantly higher than the O group on all trials. Here, however, the +NSRT group was significantly lower than the O group on Trial II, while it was not different on Trials I and III.

² Differences between means within the overall analyses of variance were evaluated by Duncan multiple-range tests. All differences refer to significance at the 5% ($p < .05$) level unless otherwise noted.

The other significant finding in the first analysis was a three-factor interaction ($F = 5.50$, $df = 1, 32$, $p < .05$). This interaction of Conditioning \times Direction \times Sex reveals the +PSRT males achieved significantly higher scores on positive self-references than did the +NSRT males. The +PSRT, +NSRT females achieved higher positive self-reference scores than did the +NSRT, +PSRT females. The +PSRT, +NSRT and +NSRT, +PSRT males and the +PSRT and +NSRT females were not significantly different. It appears from this finding that the simple conditioning situation (in the true direction only) was more effective for males than for females. The more complex conditioning situation (T&F where the subject is reinforced in the true direction for one response class and the false direction for the other response class) was more effective for females than for males.

Analysis of the Ratings to Negative Self-Reference Items

For the analysis of the ratings to negative self-references, the scores are derived exactly as in the analysis to positive self-references above. Once again the primary focus is on the conditioning variable (+PSRT versus +NSRT). Conditioning yielded an F of 4.25 representing a significant ($p < .05$) main effect which in this case showed those subjects reinforced for higher endorsements of negative self-references achieving a higher sum of the ratings on negative self-references than did those subjects who were reinforced for

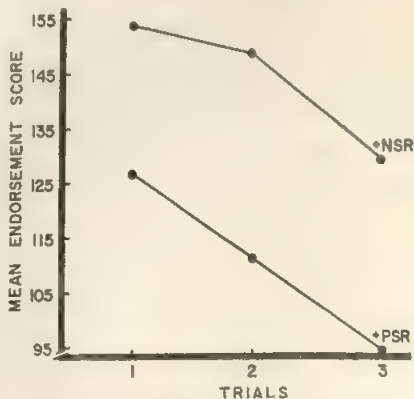


FIG. 3. Mean endorsement scores to negative self-reference items for +PSRT and +NSRT groups.

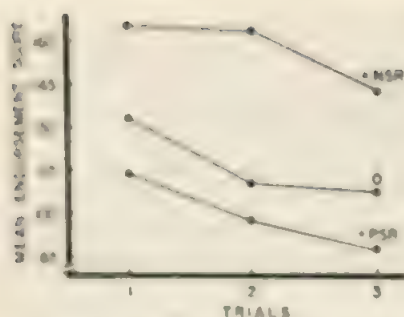


FIG. 4. Mean endorsement scores to negative self-reference items for +PSRT, +NSRT, and control groups.

higher endorsements to positive self-references (+NSR \bar{X} = 431.85, +PSR \bar{X} = 330.80).

It is clear that reinforcing subjects for higher endorsements of negative self-references resulted in higher ratings than those subjects who were reinforced for positive self-references. As in the first analysis, these data show that conditioning affected verbal behavior in the expected direction.

The examination of the trial-by-trial ratings of the negative self-reference data of +NSR versus +PSR groups is shown in Figure 3. +NSR subjects were significantly higher in acceptance of negative self-references than +PSR subjects on all trials, though they dropped over trials. This decrease over trials was significant for +PSR subjects from one trial to the next. Thus, +PSR subjects in Trial II were significantly lower on negative self-reference scores than +PSR subjects on Trial I, and +PSR subjects in Trial III were lower than +PSR subjects in Trial II. On the other hand, the decrease of +NSR subjects from Trials I to II was not significant, though it was from Trials II to III. This finding for +NSR subjects suggests that control over negative self-response ratings was maintained in Day 1 (Trials I and II), but was lost in Day 2 (Trial III).

The comparison of these groups to the O control is shown in Figure 4 for the T-only subjects. The +NSR group was significantly higher than O control on all trials. The +PSR group was significantly lower than the O control on Trials I and III, but just missed significance on Trial II. The O control group dropped significantly from Trial I to Trial

II, but leveled off, and Trial II and Trial III did not differ from the O group. The same comparisons for T&F subjects showed the +NSR group was significantly different from O in Trial II only, and the +PSR group did not differ from O at all.

A significant ($p < .05$) Sex \times Trials interaction is due to a marked drop in ratings to negative self-references for females from Trial I to Trial III and a less steep drop for males.

CHANGES ON PERSONALITY TESTS

The Semantic Differential

The semantic differential used in this experiment employed seven concepts rated against nine scales. The nine scales represented three factors with three polar scales each. The polar scales were selected because of their high loading on the particular factors and their low loadings on the other factors. The factors are evaluative—consisting of the good-bad, clean-dirty, and fair-unfair scales; potency—consisting of the large-small, strong-weak, and heavy-light scales; and activity—consisting of the fast-slow, active-passive, and hot-cold scales.

Each subject had six scores for each concept he rated—a score for evaluative, potency, and activity, both before and after testing. Since our interest was focused on the possible change of these scores from pre- to posttesting, the pretest scores were subtracted from the posttest scores. To remove negative numbers a constant of 10 was added to each number. The data were thus in a difference score (D score) form with each subject having three D scores for each of the seven concepts rated.

Analyses of variance of difference scores (Edwards, 1960) were used as the statistical tests. For each concept, three analyses were done, one for the evaluative factor, one for the potency factor, and a third for the activity factor. Each of these analyses compared the conditioning groups (+PSR versus +NSR—two levels), the direction of conditioning (T versus T&F—two levels), and the sex of the subject (two levels). They are, then, $2 \times 2 \times 2$ analyses of variance of D scores. Where significance appeared in these analyses further analyses were done comparing these groups to the neutral control groups (O), once again

rating the subjects on direction and analyzing T groups and T&F groups separately with respect to the control group.

Myself

The analysis of the ratings on the concept "Myself" revealed a significant interaction of Conditioning \times Direction for the activity factor ($p < .05$). This interaction appears due to the +NSRT, +PSRF subjects raising the activity level of the concept Myself from pre- to posttesting, while the other groups remained fairly stationary. A Duncan multiple-range test showed the +NSRT, +PSRF subjects to have significantly (see Footnote 2) higher *D* scores than the +NSRT subjects.

On the evaluative factor the ratings showed an insignificant though interesting Conditioning \times Direction \times Sex interaction ($p < .10$) in which the females in the T condition responded as was predicted. After reinforcement for positive self-references they showed higher evaluations of themselves. After being reinforced for negative self-references they showed a lower evaluation of themselves. The males under the T condition showed no differences. The same is true of males and females under the T&F condition.

Death

The concept "Death" was employed simply as a concept which has fairly universal negative connotation. The question here was what might happen to a negative, non-self-concept as a result of conditioning positive and negative self-statements. No differences appeared for either the evaluative or potency factors.

On the activity factor the Conditioning \times Direction interaction was significant at the .005 level. This interaction showed the +PSRT group going down from pre- to posttesting, while the +NSRT group rose in the rating of the activity of Death. A Duncan's analysis showed that these two groups were significantly different. In addition the +PSRT group differed from the +PSRT, +NSRF group. Comparison with O control showed the +PSRT group did have significantly lower *D* scores than the O group, but that the +NSRT group did not have significantly higher *D* scores than the O group.

The Experimenter

With regard to the concept "The Experimenter," the evaluative and activity factors showed no significant differences. All groups went down on the evaluative factor from pre- to posttesting, and all groups went up on the activity factor from pre- to posttesting except the +PSRT females, who went down on activity rating of The Experimenter.

The potency factor, on the other hand, showed a significant Conditioning \times Direction interaction ($< .025$). This interaction showed the +PSRT group and the +NSRT, +PSRT group decreasing in the potency rating of The Experimenter, while the +PSRT, +NSRF group and the +NSRT group increased in this rating. A Duncan's test revealed that the +PSRT group was significantly different than the +PSRT, +NSRF group. None of these groups were significantly different from the O control group. The control group, however, did go up in The Experimenter potency ratings.

The Good Me

For the concept "The Good Me" the evaluative and potency factors showed no significant differences. The activity factor showed a significant ($< .025$) Conditioning \times Direction \times Sex interaction. The only group to decrease in the activity rating of the The Good Me was the +NSRT males. This group differed significantly from the +NSRT females and from the +NSRT, +PSRT males. They did not, however, differ from the O males.

The Bad Me

"The Bad Me" showed no significant differences on the evaluative factor. The conditioning variable, however, showed an interesting though nonsignificant difference ($F = 3.119$, $p < .10$). A comparison with the O control subjects showed that the +PSRT group decreased in the evaluative rating of The Bad Me significantly from the O group which rose in the evaluative ratings of this concept.

The potency factor showed a main effect for direction ($F = 7.284$, $p < .025$) with those subjects conditioned T only raising the po-

factor of this concept and the subjects concluded that lowering this rating. The activity factor showed no significant differences.

The concepts, "Me—As I Would Like to Be," and "Me—As Others See me," showed no significant differences. This appeared to result from the fact that both of these were rated as being near ceiling on the pretest and remained very high in the posttesting.

Q-SORT RESULTS

In the Q-sort test, the subject had to sort 50 cards, 25 positive self-references and 25 negative self-references, into nine possible stacks. These stacks of cards ranged from the 1 end, "least like me," to the 9 end, "most like me." In addition only a certain number of cards could be put into each category, the result being a forced normal distribution. The dependent measure was that used by Rogers (1970) and others, and is called the Q Sort Adjustment Score (QSAS). Analyses were performed on the difference score ($DQSAS$) between pre- and post-QSAS scores and on the pre- and postscores themselves.

The only significant finding in the $DQSAS$ was a main factor for sex of the subject ($< .05$) which showed the males increased the QSAS score significantly more than females from pre- to posttesting. The pre- and post-QSAS scores showed no differences whatsoever. The sex difference is accountable to a slight difference between males and females before experimental procedures were initiated with the females having higher ratings on positive self-reference than the males. On the posttest the males moved up to where they were equal or slightly above the females on positive self-references. The significant D scores are then a function of the movement of males in a "healthier" direction as a result of experimental procedure, regardless of the nature of this procedure.

RESULTS OF PERSONALITY TESTS

Five personality scales were administered before and after the experimental procedure. They are the TAS, GA, H, LP, and D scales. The preconditioning score for each subject on each test was subtracted from the postconditioning score. Thus, there is a D score on each test for each subject. Analyses of variance of

the D score were run on each scale as well as analyses on the pre- and postexperimental scores. The scales are discussed separately.

Test Anxiety Scale

The TAS showed a significant Conditioning \times Direction interaction ($< .05$) and a Conditioning \times Direction \times Sex interaction ($< .05$). The first interaction appeared to be due to the +NSRT, +PSRF group which rose in TAS score from pre- to posttesting, while all other groups dropped with the +PSRT, +NSRF dropping most sharply. The Conditioning \times Sex \times Direction interaction was also due to this group, but with the addition of the females, thus the +NSRT, +PSRF females rose sharply while the remaining groups dropped or stayed the same. This group is significantly different from the +PSRT, +NSRF females. No differences were found in either the pre- or postanalyses alone.

GA showed no differences on any of the measures.

Hostility

There was an initial pretesting sex difference ($F = 9.20$, $df = 1,32$, $p < .005$) with the females attaining higher H scores than males. This same difference continued into the posttest, but was no longer significant. The post-H scores showed a significant conditioning effect ($F = 5.828$, $df = 1,32$, $p < .025$). This difference showed the +PSR subjects attaining significantly higher H scores in the posttest than the +NSR subjects. This difference existed in the pretest, but was not large enough to gain significance; thus, the D score also failed to achieve significance. Comparison with the O control showed no significant difference for the +PSR or +NSR group.

Lack of Protection Scale

This scale showed no differences in D score and pretest score, but did show a significant Conditioning \times Sex interaction in the posttest analysis ($F = 4.78$, $df = 1,32$, $p < .05$). This posttest difference was accounted for by an original difference in the pretest which did not reach significance. The +NSR males and females were the same, but the +PSR males had much lower LP scores than did the +PSR

male. This is prior to conditioning. All groups remained the same in the posttest, and the +PSR subjects dropped slightly causing the significant interaction.

Defensiveness

Defensiveness showed no significance on either pre- or *D* analyses, but did show a Direction \times Sex interaction in the posttest ($F = 6.33$, $df = 1,32$, $p < .025$). This interaction showed the male T&F group and the female T-only group higher than the other two groups. This was true in the pretest also. The interaction, then, may be a function of differences before the experimental session.

Awareness Interview

None of the subjects were able to verbalize the contingencies of reinforcement and positive or negative self-references. Three of the subjects did notice that the experimenter said "mm-hmm," but did not know why he did this.

DISCUSSION

A number of recent studies have attempted to show that changes on personality measures may result from verbal conditioning. Findings in this area, however, have not been consistent. There is a possibility that this is due to the procedure of reinforcing the frequency of the response as the relevant dimension of the response to be manipulated. For most types of research this appears to be the appropriate dimension. However, for attempts at demonstrating changes in personality measures, this may not be the case.

This study was undertaken with the view that reinforcing the frequency of the response class may have no effect on the criterion personality measures. It proposed a conditioning technique which attempted to change the verbal repertoire by manipulating the degree to which the subject endorsed a particular statement reflecting feelings about himself. The first question to be discussed is: Can verbal behavior, as reflected in degree of endorsement, be manipulated using this conditioning procedure?

The data presented support the conclusion that the technique did enable the experimenter

to control the subjects' verbal behavior. Looking at overall conditioning, that is, summed across all trials for Days 1 and 2, the subjects conditioned for positive self-references had significantly higher endorsement scores, reflecting higher endorsement levels for the positive self-reference items than did those subjects who were reinforced for negative self-references. When these two groups were compared on the negative self-reference items, the reverse was found. The two hypotheses related to the differences between the +PSR and the +NSR groups were, thus, strongly supported by the data.

A clearer demonstration of the effects of reinforcement was given by the data divided to give three matched trials. It will be remembered that in this form each of the three trials was composed of the same 120 items. Thus, changes from trial to trial represented changes in responses to the same items. These three matched trials showed a general rise for all groups on the positive self-reference items with this rise for the +PSR group significantly higher at the individual trials from the +NSR and control groups. The +PSR group also rose significantly for each succeeding trial, while the +NSR group did not rise from Trial I to II, but did rise from II to III. For the negative self-reference responses all groups dropped. Conditioning had its effect by retarding the drop of the +NSR subjects, while the +PSR and control groups dropped significantly. The +NSR group, however, did drop in Trial III.

These findings demonstrate that change can be effected on the same items when repeated, as a result of conditioning. However, the question of why all groups moved toward greater acceptance of positive self-references and toward greater rejection of negative self-references cannot be answered from the present research. A trend has been found in free-verbalization situations which is the reverse of the trend found here; that is, positive self-references tend to extinguish and negative self-references tend to increase without reinforcement. This has been reported by both Rogers (1960) and Koenig (1962). Future research should attempt to explain these trends and their relationship to the trends found in free-verbalization situations.

Conditioning fear and rage (T&F) sexual responses, which is not used to learn sex roles for the groups being conditioned in the T&F study, also led to lowered effectiveness of the procedure for the groups conditioned in the two directions. There was some indication, however, that the groups had a three set curve toward the end of the experiment. It is possible that performance was a result of the small training procedure and time was needed for the subjects to become fully aware. This suggests that a longer conditioning session than the 15-minute may be the more effective technique.

Sex was a variable in this study since it has often been found to be of importance in personality research. The findings from this study showed interaction with direction conditioning and trials. The most important finding was that the unidirectional conditioning had a greater effect for males, while the simultaneous conditioning had a greater effect for females.

An additional point of interest was that this conditioning technique appeared to be quite potent in its effect. Subjects responded to 360 items, only 156 of which were reinforceable, thus each subject received a maximum of 156 reinforcements in a total of 3-45 minutes of conditioning, yet quite strong conditioning effects were found. The technique also has the advantage of being well suited to use as a research tool. Both the items and the conditioning levels can be varied permitting a greater degree of control than is possible with other conditioning techniques.

Up to this point what has been demonstrated is that a conditioning procedure which attempts to manipulate the degree of endorsement can indeed be used for conditioning a selected response class. The second major question of this study is: Will the changes produced by this procedure have some effect on self-report personality measures?

The semantic differential showed some of the most interesting results in terms of change as a result of conditioning. Self-ratings involving the concept "Myself" showed changes in the expected direction on the evaluative factor for the +NSR females. The activity rating of the self increased as a result of reinforcing negative self-references for the T-

only group. The concept "The Good Me" was lowered in activity from pre- to posttest as a result of conditioning higher endorsement of negative self-references, though this result was confined to females under the T- only conditioning. Males, on the other hand, increased the activity ratings of "The Good Me."

The Bad Me became worse. This may be a result of conditioning differentiating "Good" and "Bad" self by widening the gap between positive and negative self-references.

The concept "Death" was increased in activity ratings as a result of conditioning higher endorsement of negative self-references, indicating that the effects of conditioning negative self-references can generalize beyond self-references to a concept not directly related to self. The concept of "The Experimenter" did not seem to be affected by conditioning except for one group, the +PSRT, +NSRF group, which increased its rating of "The Experimenter" on the potency factor. This is probably due to the fact that this group received more reinforcements than any other group.

In contrast to the results from the semantic differential, the Q-sort test showed no significant changes as a result of conditioning. From the present study, the Q sort appears to be a very stable test and as such is not well suited to procedures which attempt to change it. As a result of this stability it is impossible to determine if Rogers' failure to get generalization is indeed a function of a poor test or of the inappropriateness of frequency conditioning.

Differences on the five personality scales were primarily due to sex differences before experimental procedures were initiated. The findings related to test anxiety, however, appeared to be directly related to conditioning. In this case, the differences could be attributed to the groups getting T&F conditioning. This effect was found to be much stronger in the females than the males, a finding supported by the fact that the simultaneous T&F conditioning was more effective with females than with males.

In general, the results of the criteria personality measures support the conclusion that changes in personality measures can be brought about through the use of an appro-

conditioning technique. The strength of these changes and their relevance to actual "personality" change are questions for future research in this area.

Verbal conditioning as it affects personality was mentioned as being of particular interest to the field of psychotherapy. Many argue that therapy is a rather poorly controlled verbal conditioning procedure. This study argues against the use of reinforcement for increasing the frequency of what is being said by the patient. The present procedure would recommend some technique where the subject no longer saying the same things which were part of the personality he brought to therapy to be changed in the first place. The techniques for doing this are not necessarily new. Interpretation is a method of telling the patient something he himself will not or cannot say. Role playing is another method for instigating new responses to old situations. Thus, the therapeutic procedures may not be new, but their application should be. Greater attempts at controlling the situation so that reinforcement is contingent upon the desired new responses are necessary. We are arguing, then, for the application of operant techniques to psychotherapeutic practices, but in a way in which old behaviors are extinguished, not increased in frequency through conditioning.

REFERENCES

- BUTLER, J. M., & HATCH, G. V. Changes in the relation between self-concepts and ideal concepts contingent upon self-control training. In C. R. Franks & E. E. Thomas (Eds.), *Psychotherapy and personality change*. Chicago: University of Chicago Press, 1961. Pp. 27-37.
- FERGUSON, A. The social structure variable in personality change and control. New York: Free Press, 1967.
- FERGUSON, A. *Behavioral change in psychological research*. New York: New York: Brunner, 1966.
- KOHN, K. The manipulation of personality test scores by the technique of selective verbal reinforcement in a free verbal situation. Unpublished master's thesis, University of Washington, 1963.
- KRATZER, L. Studies of the conditioning of verbal behavior. *Psychological Papers*, 1959, 55, 115-120.
- NEUMANN, A. Conditioning of a response class on a personality test. *Journal of Abnormal and Social Psychology*, 1957, 54, 17-22.
- OSGOLD, C. E., SMITH, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Chicago: University of Illinois Press, 1957.
- ROGERS, J. M. Operant conditioning in a psychotherapy setting. *Journal of Abnormal and Social Psychology*, 1960, 60, 37-50.
- SATZINGER, K. Experimental manipulation of verbal behavior. *Journal of General Psychology*, 1959, 61, 32-44.
- SARASON, I. G. Interrelationships among individual difference variables, behavior in psychotherapy, and verbal conditioning. *Journal of Abnormal and Social Psychology*, 1958, 56, 337-344.

(Received April 12, 1966)

SELF-IMAGE DISPARITY: A DEVELOPMENTAL APPROACH¹

PHYLLIS KATZ

AND

EDWARD ZIGLER

New York University

Yale University

The present study assessed real self, ideal self, and social-self perceptions of 5th, 8th, and 11th grade children. The major prediction was that self image disparity is a function of developmental level. This hypothesis was based upon 2 factors thought to concomitantly increase with maturity: capacity for guilt and ability for cognitive differentiation. A subsidiary prediction was that measuring instruments most sensitive to the assessment of these factors should maximally reflect developmental changes in self image disparity. Both the major and subsidiary predictions received experimental support. Self-image disparity was found to be positively related to chronological age and intelligence. This larger disparity in older and brighter children was accounted for by both decreased self-evaluations and increased ideal-self images.

A frequently used measure in many empirical and theoretical contexts (cf. Wiley, 1961) has been the disparity between the individual's real-self image and ideal self. At least three different interpretations of this real-ideal-self disparity have been advanced. The most widely noted position is that of Rogers and his co-workers (Rogers & Dymond, 1954) who view such a disparity as a general indicator of maladjustment. Evidence of the popularity of this view can be seen in the practice of employing self-image disparity as an operational measure of mental illness (Scott, 1958).

A second position represents a qualification of the Rogerian thesis. A number of investigators have advanced evidence that while a large self-ideal disparity is invariably ominous, it would be found only among individuals employing particular psychological defenses, for example, sensitizers and psychoneurotics (e.g., Altrocchi, Parsons, & Dickoff, 1960; Hillson & Worchel, 1958). Within this position, maladjusted individuals employing other modes of defense, for example, denial, would be expected to show little self-ideal discrepancy.

¹This investigation was supported by Grants M-6369 and MH-06809 from the National Institute of Mental Health, United States Public Health Service, and by a New York University Arts and Science research grant. The authors wish to thank Steven Gardner for his assistance in running the children and William Bacci, superintendent of the Carl Place, New York, school system for his aid in obtaining subjects.

A third position was recently advanced by Achenbach and Zigler (1963) who employed developmental theory to generate the prediction that real-ideal-self discrepancy was positively related to the individual's level of maturity. Using the social competence index devised by Phillips and Zigler (1961) as their measure of maturity, these investigators did find a greater disparity in more mature than in less mature individuals. Contrary to Rogerian theory, the magnitude of the disparity was unrelated to the level of the individual's psychological adjustment.

The developmental rationale underlying the predicted positive relationship between self-image disparity and maturity level was based upon two factors. The first of these was that the higher the maturity level, the greater the individual's capacity for incorporating social demands, mores, and values. The high developmental person, then, makes greater self-demands, is more often unable to fulfill them, and consequently experiences more guilt than the low developmental person (Phillips & Rabinovitch, 1958; Phillips & Zigler, 1961). The second factor is based upon the work of Werner and Piaget who have discovered a greater degree of cognitive differentiation at higher levels of development. In any cognition, the more mature individual should employ more categories and make finer distinctions within each category than a less mature individual. This greater differentiating ability should result in a greater disparity when an

individual first judges his real self and then his ideal self.

This view that self-ideal discrepancy is a product of both social guilt and cognitive differentiation was tested (Achenbach & Zigler, 1963) by employing a number of measures which differentially tapped these two factors. The disparity between real self and social self, that is, the self as one believes others see it (Brownfain, 1952), was assumed to be less influenced by social guilt and did indeed result in a smaller disparity than obtained with the real-ideal-self instruments. Furthermore, the assessment of these disparities by means of an adjective check list, which allows only a "yes" or "no," and is thus less sensitive to the effects of the cognitive differentiation factor, resulted in smaller disparity scores than obtained with a questionnaire format involving six alternatives to each item. Consistent with their developmental position, Achenbach and Zigler found that the differences between high and low maturity groups in self-image disparity were greatest on instruments involving both the guilt and differentiation factors (i.e., real-ideal-self questionnaire), smaller on instruments involving a single factor (real-ideal check list and real-social questionnaire), and nonexistent on a measure in which both factors were minimized (i.e., real-social check list).

Although the findings of the Achenbach and Zigler study were in keeping with developmental thinking, it should be noted that all of the subjects employed were adults whose developmental level was assessed by a social competence index consisting of the variables of education, intelligence, employment history, occupation, and marital status. Some exception could be taken to the assumption that such an index is a very sensitive indicator of developmental level. A more direct test of the developmental position suggests itself. This position clearly generates the view that there is an ontogenetic sequence in the development of self-image disparity. A major purpose of the present study, then, was to test the specific prediction that younger children exhibit less disparity between perceived and ideal self than do older children.

Although chronological age is perhaps the most frequently used indicator of develop-

ment, there is little question that chronological age is not the most sensitive reflector of the cognitive structures which are assumed to be changing with various developmental levels or stages (Piaget, 1932; Zigler, 1963). A better indicator of the child's cognitive structure, and thus his developmental level, is his mental age. In terms of the design of the present study, then, we would not only expect more self-image disparity with increasing age, but also more disparity in high- than in low-IQ children within each age level.

As in the earlier study, six measures were collected on every subject—three image measures (real self, ideal self, and social self), each in two formats (questionnaire and adjective check list). The general expectation here was that the disparity would be highest where both guilt and cognitive differentiation were reflected, that is, real-ideal-self disparity on the questionnaire, and lowest where the effects of both these factors were minimized, that is, real-social-self disparity on the adjective check list. Furthermore, the developmental position would generate the prediction that the differences between developmental groups, defined by age and/or IQ, on self-image disparity should be greatest on measures sensitive to both factors, less on measures sensitive to a single factor, and least on measures relatively insensitive to both factors.

An ambiguity in the Achenbach and Zigler study was the failure to delineate the exact nature of the larger self-image disparity found in the more mature subjects. It would be interesting to discover whether a larger self-ideal discrepancy in more mature individuals was due to a lower self-evaluation, a higher self-expectation, or some particular combination of the two. The present study attempted to obtain data on this issue by assessing not only the amount of disparity, but these particular features of the disparity as well.

METHOD

Subjects

One-hundred-twenty children were randomly selected from all fifth-, eighth-, and eleventh-grade classes of the Carl Place, Long Island, public school system. The particular grades were chosen to include elementary school and junior and senior high school

children. Since the questionnaires used were self-administering and dependent on the subject's ability to read, it was not deemed feasible to employ subjects younger than 10 years of age (fifth grade).

Within each grade level, subjects were dichotomized on the basis of Otis Quick-Scoring Mental Ability Test scores into high and low groups. The mean IQs of the low groups were 91, 93, and 94 for the fifth, eighth, and eleventh grades, respectively. The mean IQs of the high groups were 114, 119, and 121 for the fifth, eighth, and eleventh grades, respectively. An analysis of variance of male and female subjects was conducted within each grade and intelligence classification. Thus, a three-way factorial design (Age \times IQ \times Sex) was employed which consisted of 12 groups of 10 subjects each.

Although a number of earlier studies (reviewed by Wiley, 1961) and a more recent one (McDonald & Gendler, 1963) have indicated no clear relationship between self-image disparity and socioeconomic status, use of the IQ variable sensitized the authors to the social class issue. For this reason a middle-class community was chosen that was very homogeneous in respect to socioeconomic status.

Self-Image-Disparity Measures

Two self-image disparity measures were employed. These were designed to be as similar as possible to those employed in the Achenbach-Zigler study and still be relevant for children. Some of the items used were taken from Coopersmith's (1959) scale of self-esteem. The first instrument was a questionnaire composed of 20 statements. The subject's task was to select one of six alternatives for each statement ranging from "very true" to "very untrue." Half of the statements were negative, for example, "I often wish I were someone else" and half were positive, for example, "I'm popular with the other kids." The same 20 statements were used in different orders when assessing real self, ideal self, and social self. When measuring real self, response alternatives were phrased "This is very true of me," etc.; for ideal self, "I would like this to be very true of me"; finally, for social self, "People think this is very true of me." The most positive alternative was scored with 1 point ("very true" on positive items and "very untrue" on negative items), whereas the most negative received a score of 6. Thus, the possible range for each self-measure was from 20 (most positive) to 120 (most negative).

The second instrument consisted of a list of 20 adjectives, 10 positive, for example, "successful," and 10 negative, for example, "sneaky." For each adjective there were only two response alternatives, "yes" or "no." Analogous to the questionnaire, three different orders of the list were given to assess real self, ideal self, and social self. Positive responses were scored 0 ("no" to negative or "yes" to positive adjectives) and negative responses scored 1. The possible range on this instrument was from 0 (most positive) to 20 (most negative).

Considerable evidence has now been presented that the type of instruments used in this study are

reliable in the chronological and mental age groups employed in this study (Engel, 1959; McAfee & Cleland, 1965; Perkins & Shannon, 1965).

Procedure

The instruments were group administered to entire classes in their schoolrooms. The subjects were told that the experimenter wanted to find out what children thought about themselves. They were told that there were no right or wrong answers, that their responses would be kept in strict confidence, and that nobody at the school would be permitted to see the questionnaires. Specific instructions as to how to respond were printed at the top of each section, and a sample item was demonstrated by the experimenter on the blackboard. The questionnaires and adjective lists were always administered in the same order, namely, real self, ideal self, and social self with the three questionnaires first and the three adjective lists next.

RESULTS

The mean group absolute and disparity scores obtained on both the questionnaire and the adjective check list are contained in Table 1.

Disparity Between Real and Ideal Self

A Lindquist (1953) repeated-measures analysis of variance (Measure \times Age \times IQ \times Sex) was conducted on the real- and ideal-self questionnaire scores. The findings relevant to the hypotheses of this study are contained in the within-subjects portion of this analysis. These findings revealed a significant measure effect ($F_{1,108} = 267.64, p < .001$), a significant Measure \times Age interaction ($F_{2,108} = 11.60, p < .001$), and a significant Measure \times IQ effect ($F_{1,108} = 11.12, p < .01$). As can be seen in Table 1, real-self scores are greater (more negative) than ideal-self scores. The two significant interactions indicate that the magnitude of this difference between real and ideal scores is influenced by both age and IQ. The Measure \times Age interaction reflects the greater differences between real and ideal scores at Grades 11 and 8 than at Grade 5. The mean differences at these grades were 23.0, 22.1, and 11.1, respectively. The Measure \times IQ interaction reflects the finding that the high-IQ subjects exhibit a greater discrepancy between real and ideal scores than do low-IQ subjects (22.3 versus 15.1).

In order to further assess these differential discrepancy scores, real-self and ideal-self

TABLE 1
SELF-IMAGE SCORES FOR EACH GROUP

Grade	N	Questionnaire					Adjective check list				
		Absolute scores			Disparity		Absolute scores			Disparity	
		Real	Ideal	Social	Real-ideal	Real-social	Real	Ideal	Social	Real-ideal	Real-social
5th grade											
Low IQ	20	57.4	46.4	55.6	11.0	2.2	4.2	1.9	4.8	2.3	-.6
High IQ	20	51.2	40.0	49.0	11.2	2.2	3.2	.9	3.0	2.3	.2
8th grade											
Low IQ	20	61.6	46.2	54.2	15.4	7.4	5.5	1.0	4.3	4.5	1.2
High IQ	20	60.4	31.6	57.1	28.8	3.3	6.6	1.0	5.4	5.6	.8
11th grade											
Low IQ	20	57.4	38.4	51.4	19.0	6.0	4.6	.5	4.4	4.1	.2
High IQ	20	60.8	33.9	54.7	26.9	6.1	6.5	.6	5.4	5.9	1.1

questionnaire scores were analyzed separately in Age \times IQ \times Sex analyses of variance. The analysis of real-self scores revealed a significant difference associated with age ($F_{2,108} = 5.36$, $p < .01$). The real-self ratings were found to be more negative at Grades 11 and 8 than at 5 (59.1, 61.0 versus 54.3). A significant Age \times IQ interaction ($F_{2,108} = 3.06$, $p < .05$) was also found. As can be seen in Table 1, this interaction reflected the different relative positions of high- and low-IQ children at the three grade levels. In the fifth grade, high-IQ subjects express more positive feelings about themselves than the low-IQ subjects; at the eighth grade, high- and low-IQ subjects are similar; and at the eleventh grade, high-IQ subjects express more negative real-self ratings than their less intelligent peers. On the ideal-self analysis, significant age ($F_{2,108} = 4.36$, $p < .05$) and IQ ($F_{1,108} = 18.77$, $p < .001$) main effects were found, indicating that the older or brighter the child, the more positive was his ideal self.

The repeated-measures analysis of variance of the real- and ideal-self scores on the adjective check list resulted in findings similar to those obtained on the questionnaire instrument. A significant measure effect ($F_{1,108} = 253.33$, $p < .001$) and a significant Measure \times Age interaction ($F_{2,108} = 12.69$, $p < .001$) were found, reflecting greater negative scores on the real than on the ideal measure, and an increase in real-ideal disparity with age. A Measure \times IQ interaction of borderline sig-

nificance ($p < .10$), was also found. This trend was in the same direction as the significant findings on the questionnaire and reflects the tendency for greater real-ideal disparity in high- than in low-IQ subjects.

Analyses conducted separately on the real- and on the ideal-self adjective scores revealed only significant age effects on both the real ($F_{2,108} = 7.09$, $p < .01$) and ideal ($F_{2,108} = 4.22$, $p < .05$) scores. With increasing age, the real scores became more negative, and the ideal scores more positive.

Disparity between Real and Social Self

The repeated-measures analysis done on the real and ideal scores was also run on the real and social scores. On the real and social questionnaire scores this analysis resulted in a significant measure effect ($F_{1,108} = 45.30$, $p < .001$) and a significant Measure \times Age interaction ($F_{2,108} = 3.93$, $p < .05$). As can be seen in Table 1, the social-self scores were smaller (more positive) than the real-self scores. The interaction indicates that the magnitude of difference between real- and social-self scores increases with age. Unlike the real-ideal discrepancy, however, the real-social discrepancy was not significantly influenced by IQ.

An Age \times IQ \times Sex analysis of variance conducted on the social questionnaire scores alone revealed no significant effects.

The repeated-measures analysis conducted on the real and social adjective check-list

scores revealed a significant measure effect ($F_{1,108} = 10.59, p < .01$) and a significant Measure \times Age interaction ($F_{2,108} = 6.80, p < .01$), similar to those found on the questionnaire instruments. Social-self scores were smaller (more positive) than real-self scores, and the difference between the two scores was greater at Grades 8 and 11 than at 5. In addition, a significant Measure \times Age \times Sex interaction ($F_{2,108} = 4.41, p < .05$) and a significant Measure \times Age \times Sex \times IQ interaction ($F_{2,108} = 3.24, p < .05$) were found. These interactions primarily reflect the finding that fifth-grade male subjects perceived others evaluating them more negatively than they evaluated themselves. This trend was particularly pronounced in the low-IQ subjects at this age level. All other groups exhibited a real-social disparity in the opposite direction.

An Age \times IQ \times Sex analysis of variance conducted separately on the social adjective check-list scores revealed a significant Age \times IQ interaction ($F_{2,108} = 3.18, p < .05$) and a significant Age \times Sex interaction ($F_{2,108} = 5.59, p < .01$). The Age \times IQ interaction reflects the fact that the social scores are greater for low- than for high-IQ subjects at Grade 5, whereas an opposite pattern is exhibited at Grades 8 and 11. The Age \times Sex interaction can be described as follows: At the fifth grade the boys see themselves being evaluated more negatively than girls; at the eighth grade, girls see themselves being evaluated more

negatively than boys; finally, at the eleventh grade, boys and girls see themselves evaluated approximately the same.

Comparisons of Self-Image Disparity across the Different Measures

The findings reported to this point are generally in keeping with the major hypothesis that the magnitude of self-image disparity is related to developmental factors. A subsidiary hypothesis was that this positive relationship between development and self-image disparity is due specifically to two factors—an increase in guilt, and an ability to make finer cognitive judgments. A test of this hypothesis requires comparisons of self-image disparity on instruments differing in their susceptibility to the effects of these two factors. The expectation here was that the greatest disparity should be found on the instrument most sensitive to both factors (the real-ideal questionnaire), and the smallest disparity on the instrument least sensitive to both factors (the real-social adjective check list). Furthermore, it is the most sensitive instrument that should maximally reflect the developmental variables of age and IQ.

It is clearly inappropriate to test this hypothesis by comparing the discrepancy scores obtained on the questionnaires (a maximum possible score of 100) with those obtained on the check list (a maximum possible score of 20). In order to make all the discrepancy scores comparable, the following analyses were conducted on the total number of items changed between the two measures, ignoring magnitude of change. It should be noted that a discrepancy score defined by frequency of change was the same as that employed by Achenbach and Zigler (1963). Table 2 presents these discrepancy scores for each group.

These scores were subjected to a Measure (real-ideal versus real-social) \times Instrument (questionnaire versus check list) \times Age \times IQ repeated-measures analysis of variance. With the expectation of the higher than expected scores obtained on the real-social check list, the findings of this analysis were consistent with the hypotheses advanced. The main effect of measure was significant ($F_{1,114} = 14.01, p < .001$), reflecting the finding that real-ideal disparity was greater than the real-

TABLE 2
DISPARITY SCORES DEFINED BY NUMBER OF CHANGES

Group	Real-ideal-self disparity		Real-social-self disparity	
	Questionnaire	Check list	Questionnaire	Check list
5th grade				
Low IQ	11.00	5.10	10.80	5.60
High IQ	10.10	3.90	9.25	4.65
8th grade				
Low IQ	13.50	6.35	11.90	7.50
High IQ	15.20	6.90	10.55	6.70
11th grade				
Low IQ	12.55	4.80	11.30	6.80
High IQ	16.50	6.70	11.50	7.05
Total	78.85	33.45	64.80	38.30

social disparity across the two instruments. This would be predicted from the hypothesis since the real-ideal questionnaire is assumed to be sensitive to two factors, the real-ideal check list to one factor, and the real-social check list to no factors. Similarly, a significant instrument effect ($F_{1,114} = 401.41, p < .001$) was obtained reflecting the greater total discrepancy scores on the questionnaires (sensitive to two and one factor) than on the check lists (sensitive to one and no factors). A significant Measure \times Instrument interaction ($F_{1,114} = 113.42, p < .001$) was found. This is a complex interaction that reflects a number of findings congruent with the predictions and one finding antithetical to the prediction. The discrepancy scores on the self-ideal questionnaire are significantly larger ($t = 5.21, p < .001$) than those made on the self-social questionnaire (predicted), whereas the discrepancy scores on the self-ideal check list tend to be smaller ($t = 1.85, p < .01$) than those on the self-social check list (opposite to prediction). The discrepancy scores on the self-ideal questionnaire were greater ($t = 17.16, p < .001$) than those on the self-ideal check list (predicted). Finally, the self-social questionnaire disparity is greater ($t = 11.11, p < .001$) than the self-social check-list disparity (predicted).

As predicted, a number of interactions between the particular disparity scores and the developmental variables were obtained. The effect of measure interacts significantly with both grade ($F_{2,114} = 4.24, p < .05$) and IQ ($F_{1,114} = 17.99, p < .001$), indicating that the difference between real-ideal and real-social scores is greater with increasing age and higher IQ. A significant Measure \times Grade \times IQ interaction ($F_{2,114} = 3.63, p < .05$) was found. Although a number of differences contribute to this interaction, a sizable portion of it would appear to be due to the finding that the greatest difference between high- and low-IQ groups was obtained with real-ideal disparities at the eleventh grade. A predicted Measure \times Instrument \times Grade interaction ($F_{2,114} = 11.37, p < .001$) was also found. This interaction primarily reflects the greater sensitivity of the self-ideal questionnaire disparity to grade differences. This is especially noticeable between the fifth and eighth grades,

where the difference on this instrument is considerably greater than that found with the other three instruments.

The final significant finding revealed by this analysis was an Instrument \times Measure \times IQ interaction ($F_{1,114} = 7.65, p < .01$). As can be seen in Table 2, this interaction reflects the following: on the two real-ideal measures, high-IQ subjects have greater disparity scores than low-IQ subjects with the reverse being true for the two real-social measures, and the magnitude of this crossover effect is more pronounced on the questionnaire instruments than on the check-list instruments.

Further evidence that the instrument (real-ideal questionnaire) sensitive to both factors is most sensitive to the developmental variables of grade and IQ can be obtained directly from Table 2. It is with this measure of disparity that we obtain the greatest differences between grades and IQ levels. The only finding in Table 2 contrary to prediction is the larger discrepancy scores found in the real-social check list as compared to the real-ideal check list. However, the question of the comparability of these two scores must be raised. A phenomenon not encountered in the earlier study with adults (Achenbach & Zigler, 1963) must be noted. Whereas the disparity on the real-ideal check list was typically in the expected direction, with the ideal being more socially desirable than the real, this was less true on the real-social check list. Unlike adults, the children were quite willing to report that they were seen by their peers as being less socially desirable on certain traits than they had reported themselves as being on the real-self check list. The failure to take into consideration the direction of the discrepancy in the scores reported in Table 2 may have led to an erroneous indicator of what we are primarily concerned with, namely, the disparity score predicated upon a change to a more positive social self. In order to investigate this possibility, new disparity scores were calculated for the real-ideal and real-social check-list data which included only the number of items in which the disparity indicated a change from negative to positive. These scores are presented in Table 3.

A comparison of these scores with those reported in Table 2 supports the hypothesis

TABLE 3

FREQUENCY OF NEGATIVE TO POSITIVE CHANGES
IN ADJECTIVE CHECK LIST

Group	Disparity	
	Real-ideal	Real-social
5th grade		
Low IQ	3.65	2.50
High IQ	3.10	2.35
8th grade		
Low IQ	5.35	4.35
High IQ	6.20	3.95
11th grade		
Low IQ	4.45	3.55
High IQ	6.15	4.20
Total	28.90	20.90

that the larger disparity scores in the real-social check list reported in Table 2 were due to the large number of plus to minus instances. A Grade \times IQ \times Measure analysis of variance of the data reported in Table 3 revealed a significant measure effect ($F_{1,108} = 20.56, p < .001$) reflecting the predicted greater disparity on the real-ideal than on the real-social check-list scores. It thus appears that the one earlier finding that was in a direction opposite to that predicted by the two-factor developmental hypothesis was the result of ignoring directionality in the discrepancy scores. (In respect to the analysis of the data presented in Table 2, ignoring the directionality factor resulted in a more conservative test of the hypotheses under test, which nevertheless received substantial statistical support.)

Other Findings

The assumption that less mature subjects had a less differentiated response tendency than more mature subjects was tested further by running a repeated-measures analysis (Age \times IQ \times Sex \times Measure) on the frequency with which each subject utilized the extreme response categories of "very true" and "very untrue" on each of the three questionnaires. These scores are presented in Table 4. The significant main and interaction effects found generally support the cognitive differentiation hypothesis. The children were found to make

fewer extreme responses as they got older ($F_{2,108} = 3.75, p < .05$) with the brighter children giving fewer extreme responses than the less intelligent ones ($F_{1,108} = 6.40, p < .05$). As might be expected, ideal-self ratings evoked more extreme responses than either the real-self or social-self ratings ($F_{2,216} = 126.80, p < .001$). The significant Measure \times Age ($F_{4,216} = 4.30, p < .01$) and Measure \times IQ ($F_{2,216} = 9.10, p < .001$) interactions reflect the relative insensitivity of the ideal-self measure to the effects of age and IQ. It is on the real- and social-self measures that one finds a decrease in extreme responses with age and/or higher intelligence. A significant IQ \times Sex \times Measure interaction ($F_{2,216} = 3.65, p < .05$) was also found reflecting the one exception to the general findings. On the ideal-self ratings, and only on the ideal-self ratings, high-IQ boys were found to make more extreme responses than low-IQ boys. This may be accounted for by a higher level of aspiration on the part of brighter males, since extreme scores on ideal-self ratings may indicate perfectionistic strivings. It is interesting to note, however, that this exception did not occur in females.

In order to assess the construct validity of the self-image-disparity concept, a correlation was computed between each subject's real-ideal disparity on the questionnaire and that obtained on the check list. The degree of relationship was highly significant ($r = .69, p < .001$).

TABLE 4

MEAN NUMBER OF EXTREME RESPONSES
TO QUESTIONNAIRE

Group	Measure		
	Real self	Ideal self	Social self
5th grade			
Low IQ	8.8	13.3	11.6
High IQ	7.7	12.0	7.0
8th grade			
Low IQ	7.0	12.3	7.4
High IQ	5.4	11.9	5.0
11th grade			
Low IQ	6.5	11.8	7.7
High IQ	4.2	12.2	4.4

DISCUSSION

The findings of the present study lend considerable credence to the view that self-image disparity increases with increasing maturity. Real-ideal-self disparity was found to be a positive function of both chronological age and IQ. This is a rather surprising finding in light of the conventionally dim view that has been taken of an increasing self-image disparity and the rather negative psychodynamics that are thought to accompany it (McCandless, 1961; Rogers & Dymond, 1954). The findings are in accordance with earlier results obtained with adults of varying maturity levels (Achenbach & Zigler, 1963) and suggest that self-image disparity might be better conceptualized as an index of development rather than a measure of maladjustment.

Although such a view is an unconventional one, further support can be found in certain underemphasized findings in the literature. Coopersmith (1959), in a study of fifth and sixth graders, did find that self-ideal discrepancy was positively related to anxiety. However, he also discovered that children having the largest self-ideal discrepancies received the highest ratings by others, had the highest need-achievement scores, and the highest actual achievement. This is reminiscent of Brownfain's (1952) finding that college students having unstable self-concepts made better grades and were rated more intelligent than students having more stable self-concepts. McAfee and Cleland (1965), employing retardates having mental ages similar to those of the youngest group in the present investigation, found that self-ideal disparity was unrelated to adjustment, but was positively related to IQ. Perkins and Shannon (1965), employing a sample of sixth graders, did not find a significant relationship between real-ideal disparity and IQ. However, as in the present study they did find a positive relationship between ideal scores and IQ. As noted by Achenbach and Zigler (1963), exponents of the stylistic approach to the understanding of self-image disparity, for example, Altrocchi et al. (1960), Hillson and Worchel (1958), may well have obtained their findings (individuals with certain types of defenses have higher disparity scores than

individuals employing other types of defenses) by inadvertently comparing individuals differing in maturity levels. As noted in the earlier paper (Achenbach & Zigler, 1963), the developmental position takes as its given the level of maturity attained by the individual and sees both the defenses employed and the amount of self-image disparity as an outgrowth of this level. Within this framework, the degree of self-image disparity would be expected to be low at low levels of maturity and high at high levels of maturity. However, since one finds both adjusted and maladjusted people at all levels of maturity, no simple relationship between degree of self-image disparity and adjustment would be expected.

While there have been studies with children indicating a positive relationship between self-image disparity and both paper-and-pencil tests of adjustment (Hanlon, Hofstaetter, & O'Connor, 1954) and judgments of being "less secure" (Bruce, 1958), the most consistent finding in research on children's self-image disparity is the positive relationship typically found between self-ideal disparity and anxiety (Bruce, 1958; Coopersmith, 1959; Lipsitt, 1958).

It is this consistently found self-image-disparity-anxiety relationship that has probably led workers, for example, McCandless (1961), to emphasize the ominous nature of an increasing self-image disparity. The implicit assumption that anxiety is an essentially negative agent in the individual's total psychic economy would appear to be open to considerable dispute. There are perhaps as many instances in which anxiety is beneficial as those in which it is detrimental to the individual (Ruebush, 1963). The key determinant would appear to be not the presence or absence of anxiety, but rather the individual's response to anxiety in particular situations. This basically positive view of anxiety is consistent with Hebb's (1958) argument that the capacity for anxiety increases as one ascends the phylogenetic scale due to the increasing cognitive capacity of the organism. Provided one can apply Hebb's position to ontogenetic development, the relationship between self-image disparity and anxiety becomes quite understandable within the developmental

framework advanced in this paper. Rather than being ominous in nature, increasing self-image disparity would invariably appear to accompany the attainment of higher levels of development, since the greater cognitive differentiation found at such levels must invariably lead to a greater capacity for self-derogation, guilt, and anxiety. As Achenbach and Zigler (1963) noted, the attainment of higher developmental levels does not constitute an unmitigated blessing. While such attainment guarantees the individual a greater ability to deal with whatever problems confront him, his greater cognitive differentiation also gives him the capacity to construct more problems for himself.

Support was found for the subsidiary hypothesis that self-ideal disparity, as usually measured, is a function of two underlying factors related to maturity level, namely, capacity for guilt and cognitive differentiation. On the basis of this hypothesis, it was predicted that various instruments differing in their sensitivity to these two factors should differentially assess the magnitude of the self-ideal disparity. Thus, the real-ideal questionnaire measure (assumed to be reflecting both social guilt and cognitive differentiation) was expected to be maximally sensitive to developmental trends, the real-social-self questionnaire (cognitive differentiation only) and real-ideal adjective (social guilt only) measures to be next most sensitive, and finally the real-social adjective measure (reflecting neither factor) to be least sensitive to age and intelligence effects. For the most part, the results supported these expectations.

An additional indication of the importance of the cognitive differentiation factor was revealed in the finding that the number of extreme responses is negatively related to maturity level, defined by both age and IQ. These findings were perfectly consistent with those reported by Light, Zax, and Gardener (1965). These investigators, employing exactly the same developmental rationale concerning cognitive differentiation as that utilized in this study, found that older and brighter children made fewer extreme responses in rating Rorschach inkblots on semantic differential scales.

Another finding of interest relates to the question of just what aspect of the self-image disparity changes with development. McCandless (1961) has posited that self-ideal-discrepancy scores could be easily replaced by single real-self measures. This argument is based upon the assumption that ideal-self ratings are not ordinarily subject to individual variation. The results of the present investigation contradict this assumption. The increase in real-ideal-self disparity with age was accounted for by both significantly lowered self-evaluations and more positively defined ideal-self ratings. Thus, both aspects of the disparity are related to development. In general, the magnitude of difference was greatest between the fifth- and eighth-grade children and least between the eighth- and eleventh-grade groups, thus suggesting that early adolescence may be a pivotal point in the development of self-image disparity.

The authors have presented their argument in such a way as to highlight the differences between their position and the Rogerian view that a large self-image disparity is ominous in nature. It should be noted, however, that this latter position is a logically appealing one, and considerable evidence has been presented indicating that a large self-image disparity is often accompanied in adults by a state of malaise and maladaptive behavior (Rogers & Dymond, 1954). Perhaps a judicious conclusion of the present study would be that future investigators should be cognizant of developmental factors when interpreting self-image-disparity findings and should expend some energy in determining exactly how the psychodynamic factors emphasized by many workers interact with the developmental phenomena investigated in this paper.

REFERENCES

- ACHENBACH, T., & ZIGLER, E. Social competence and self-image disparity in psychiatric and nonpsychiatric patients. *Journal of Abnormal and Social Psychology*, 1963, 67, 197-205.
- ALTROCCHI, J., PARSONS, O. A., & DICKOFF, H. Changes in self-ideal discrepancy in repressors and sensitizers. *Journal of Abnormal and Social Psychology*, 1960, 61, 67-72.
- BROWNFAIN, J. J. Stability of the self-concept as a dimension of personality. *Journal of Abnormal and Social Psychology*, 1952, 47, 597-606.

- BRUCE, P. Relationship of self-acceptance to other variables with sixth-grade children oriented in self-understanding. *Journal of Educational Psychology*, 1958, 49, 229-238.
- COOPERSMITH, S. A method for determining types of self-esteem. *Journal of Educational Psychology*, 1959, 59, 87-94.
- EXCEL, M. The stability of the self-concept in adolescence. *Journal of Abnormal and Social Psychology*, 1959, 58, 211-215.
- HANLON, T. E., HOPSTAETTER, P. R., & O'CONNOR, J. P. Congruence of self and ideal self in relation to personality adjustment. *Journal of Consulting Psychology*, 1954, 18, 215-218.
- HEBB, D. The mammal and his environment. In C. Ree, J. Alexander, & S. Tomkins (Eds.), *Psychopathology: A source book*. Cambridge: Harvard University Press, 1958. Pp. 127-135.
- HILLSON, J. S., & WORCHEL, P. Self-concept and defensive behavior in the maladjusted. *Journal of Consulting Psychology*, 1958, 22, 45-50.
- LIGHT, C. S., ZAX, M., & GARDENER, D. H. Relationship of age, sex, and intelligence level to extreme response style. *Journal of Personality and Social Psychology*, 1965, 2, 907-909.
- LENDQUIST, E. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
- LIPSITT, L. P. A self-concept scale for children and its relationship to the children's form of the Manifest Anxiety Scale. *Child Development*, 1958, 29, 463-472.
- McAFEE, R. O., & CLELAND, C. C. The discrepancy between self-concept and ideal-self as a measure of psychological adjustment in educable mentally retarded males. *American Journal of Mental Deficiency*, 1965, 70, 63-68.
- McCAGLESS, B. R. *Children and adolescents*. New York: Holt, Rinehart & Winston, 1961.
- MCDONALD, R. L., & GYNTHER, M. D. Relationship of self and ideal self descriptions with sex, race, and class in southern adolescents. *Journal of Personality and Social Psychology*, 1965, 1, 85-88.
- PERKINS, C. W., & SHANNON, D. T. Three techniques for obtaining self-perceptions in preadolescent boys. *Journal of Personality and Social Psychology*, 1965, 2, 443-446.
- PHILLIPS, L., & RABINOVICH, M. S. Social role and patterns of symptomatic behavior. *Journal of Abnormal and Social Psychology*, 1958, 57, 181-186.
- PHILLIPS, L., & ZIGLER, E. Social competence: The action-thought parameter and vicariousness in normal and pathological behavior. *Journal of Abnormal and Social Psychology*, 1961, 63, 137-146.
- PIAGET, J. *The moral judgment of the child*. London: Routledge & Kegan Paul, 1932.
- ROGERS, C., & DYMOND, R. *Psychotherapy and personality change*. Chicago: University of Chicago Press, 1954.
- RUEBUSH, B. Anxiety. In H. W. Stevenson (Ed.), *Child psychology: The 62nd Yearbook of the National Society for the Study of Education*. Part 1. Chicago: University of Chicago Press, 1963. Pp. 460-516.
- SCOTT, W. Research definitions of mental health and mental illness. *Psychological Bulletin*, 1958, 55, 1-45.
- WILEY, R. *The self-concept*. Omaha: University of Nebraska Press, 1961.
- ZIGLER, E. Metatheoretical issues in developmental psychology. In M. Marx (Ed.), *Psychological theory*. (2nd ed.) New York: Macmillan, 1963. Pp. 341-369.

(Received April 13, 1966)

PERCEIVED INSTRUMENTALITY OF THE COLLEGE AS A MEASURE OF ATTITUDES TOWARD COLLEGE

ANNE CONSTANTINOPOLE

Brins Made College

Students' attitudes toward college were measured with a test designed to measure the extent to which college was thought to help in the attainment of important goals. The 34 goal statements, covering 11 dimensions of the college experience, were rated for the degree to which the goal was thought to be helped or hindered goal attainment. The summed product of these ratings measured instrumentality, an independent rating of satisfaction with college. All 11 dimensions were significantly related to instrumentality. High correlation with the summed level of ratings of the instrumentality of the college was found for the instrumentality of the college. In most instances, females gave higher ratings of both evaluation and instrumentality than males.

The primary purpose of the larger study, of which this paper represents a part, was an investigation of the correlates of level of happiness among college students. One important correlate seemed to be the student's perception of the college environment in terms of how well the college was meeting his needs. Although this perception is probably reflected in a generalized rating of satisfaction with the college, the work of McConnell and Heist (1962) indicates that the diversity of student goals should be considered in any attempt to assess attitudes toward college. Birney and Taylor (1959) were cognizant of this problem when they developed the Orientation to College Inventory for measuring the attitudes of students at Amherst College. They assumed both a diversity of motivations among students and a diversity of rewards within the college environment, and also assumed that the student will orient himself toward that aspect of the environment in which he obtains maximum rewards. However, their instrument included only two broad aspects of the college environment. In the present study a measure was needed which would take into account a greater variety of goals, and which would also allow for separate evaluations of their importance and of the degree to which the stu-

dent believed he was obtaining the desired rewards.

The instrument developed to meet these requirements was based on the theory of attitude structure developed by Peak (1955) and Rosenberg (1956), and was essentially a modification of Rosenberg's test of perceived instrumentality. According to Peak (1955), "attitudes as dependent variables are a function of: 1) the instrumentality of their referent objects or situations for aiding or interfering with goal attainment, and 2) the satisfaction derived from reaching goals [p. 158]." In Rosenberg's original investigation, the attitude object was "the policy of allowing members of the Communist party to address the public." Ratings were made of (a) the degree to which this policy would aid or interfere with the attainment of each of 35 values, and (b) the importance of each value to the subject. Rosenberg found that the product of these two ratings summed across all 35 values was highly related to independent prior evaluations of the policy. Carlson (1956) used a similar measure to demonstrate that, when subjects are made aware of the relevance of an attitude object to important values, changes in the perceived instrumentality of the object are associated with attitude change. In adapting this measure for the present purposes, it was assumed that (a) the ratings of perceived instrumentality would reflect the rewards which the student obtains in the college environment, and (b) the ratings of the goals

¹ This paper is based in part on a doctoral dissertation submitted to the Department of Psychology, University of Rochester. The author is indebted to Vincent Nowlis for guidance in all phases of the research.

reflect the areas in which he is most expected to succeed.

In an earlier study at the University of Rochester, Higgins (1963) had obtained ratings of the importance in the subject's college experience of each of 88 items. These items had been taken largely from measures used in earlier studies, and pertained to many different aspects of the college experience. The 754 subjects were freshman and senior men and women; their ratings were factor analyzed separately for each sex/class group. Although there were sex and class differences in mean ratings, the results revealed nine factors which appeared in all four matrices, each factor representing a separate dimension of the college experience. The majority of the statements of goals used in the present measure are adapted from the items which loaded most highly on these nine factors.

Three hypotheses were tested: (a) The summed product score from the Perceived Instrumentality of the College Test (PICT) will be positively correlated with an independent measure of satisfaction with the college; (b) happier students will perceive the University as more helpful in their progress toward important goals than will less happy students; and (c) the diversity of motivations and rewards among students will be apparent in sex and class differences in both the evaluation and instrumentality ratings of the PICT.

METHOD

Subjects

The subjects were 99 freshman males, 89 freshman females, 90 junior males, and 75 junior females. The subjects were recruited from classes in 10 academic departments in the College of Arts and Science at the University of Rochester. All were full-time undergraduate students.

Measures

The PICT consists of 14 statements, each of which might be described as a goal in the college experience. Two ratings on a 7-point scale were made for each goal: (a) its importance as a goal in the subject's own college experience, and (b) the degree to which the University was perceived as helping or hindering progress toward that goal. The product of the two ratings was obtained for each goal and the products summed across all 14 goals in order to obtain the general index of attitude toward

TABLE 1

Reliability of the PICT

Goal	Importance	Helpfulness	Product
Learning about the world	.44**	.42**	.51**
Learning about the self	.44**	.42**	.51**
Acquiring an education	.44**	.42**	.51**
Developing a career	.44**	.42**	.51**
Developing a personality	.44**	.42**	.51**
Contributing in a discipline	.46**	.39	.64**
Developing a career with different friends	.52**	.34	.50**
Becoming self-confident	.47**	.54**	.65**
Learning a spouse	.48**	.08	.42**
Achieving academic distinction	.48**	.65**	.61**
Having many good friends	.57**	.56**	.70**
Discovering your own strong points and limitations	.45**	.42**	.71**
Preparing for a career which begins right after graduation	.57**	.65**	.69**
Preparing for a career which requires further study beyond the BA or BS	.70**	.52**	.74**
Summed scores	.62**	.67**	.72**

* $p < .05$, $r = .40$, $r^2 = .40$.

** $p < .01$, $r = .49$, $r^2 = .49$.

college. The only test-retest data available for computation of the reliability of the PICT were from a group of first-semester freshmen ($N = 41$). These data probably underestimate the reliability of the instrument with other groups of college students, since so much change in attitudes occurs during the freshman year. The test-retest correlations are presented separately in Table 1 for the evaluation rating, the instrumentality rating, and the product of the two ratings for each of the 14 goals and for the summed score. The reliability for the total summed product score was .72.

The independent measure of attitude was obtained with a 7-point graphic scale on which the student was asked to indicate his degree of satisfaction or dissatisfaction with the University.

TABLE 2
MEANS, STANDARD DEVIATIONS, AND *t* TESTS
FOR THE ELATION-DEPRESSION SCALE

Sex and class	Means				<i>t</i> test
	M	SD	M	SD	
1968 males	34.73	9.17	34.83	8.93	1.02
1968 females	42.31	8.13	32.56	12.94	2.27*
1966 males	42.38	10.24	38.18	6.17	2.81*
1966 females	43.63	10.81	36.75	13.67	1.78

* $p < .05$.
* $p < .01$.

Level of happiness was measured with a 10-point bipolar scale developed by Weisman and Ricks (1966) to measure the elation-depression dimension of mood in college students. Each subject was asked to indicate which one of the 10 statements best described his average level of happiness or unhappiness during the current academic year. Test-retest reliability of this measure, with 6 weeks between administrations, was .85 ($N = 150$).

Procedure

Two questionnaires were used to obtain the data; the first was completed in the classroom the second at home. The measures of level of happiness and satisfaction with college were part of the first questionnaire, while the PICT was contained in the second. Data were gathered during the first 2 weeks of March (during the fifth and sixth weeks of a 16-week semester).

RESULTS

Perceived Instrumentality and Satisfaction

One test of the validity of the PICT involves its correlation with an independent measure of satisfaction with college. Although these two measures differ considerably in the extent to which relevant components of the attitudes are specified, there should be a positive relationship between them. The Pearson product-moment correlations of the summed-products score on the PICT with the rating of satisfaction were: freshman males, .46; freshman females, .37; junior males, .34; and junior females, .49. All correlations are significant at the .01 level.

Perceived Instrumentality and Level of Happiness

According to the second hypothesis, students who avow a relatively high average

level of happiness should perceive the University as helping in their progress toward important goals. Therefore, it was expected that their summed products score would be higher than that of the relatively unhappy students. This hypothesis was tested using the data from the 16 highest- and 16 lowest-scoring subjects on the Elation-Depression scale. Table 2 presents these data and the results of the *t* tests for each of the four sex/class groups. The differences are significant for three of the four groups.

In order to discover the relative contribution of the products for each of the 14 goals to these results, correlations were computed between level of happiness and each of the 14 products. These correlations are presented separately for each of the four sex/class groups in Table 3. For both groups of juniors and for the freshman males, the product score for the goal "Becoming self-confident" was most closely related to level of happiness. For the freshman females, level of happiness was most closely associated with "Establishing your own personal, social, and academic values." Sex differences were apparent in these data: "Preparing for a career which requires further study beyond the BA or BS," "Achieving academic distinction," and "Personal independence" were significantly related to level of happiness among the males but not the

TABLE 3
CORRELATIONS OF PRODUCT SCORE WITH
ELATION-DEPRESSION SCALE

Item	1968 males ($N = 99$)	1966 males ($N = 90$)	1968 females ($N = 89$)	1966 females ($N = 75$)
Learning	.29**	.16	.07	.00
Ideas	.25**	.05	.25**	.02
Values	.15	.31**	.44**	.17
Opposite sex	.29**	.18	.29**	.14
Campus group	.11	-.02	.29**	.21*
Get along	.19	.15	.14	.14
Confidence	.39**	.38**	.02	.32**
Independence	.24*	.28**	.17	.05
Spouse	.21*	.18	.20*	.11
Distinction	.27**	.24*	.08	.03
Friends	.36**	.17	.28**	.19
Limitations	.12	.20*	.10	.07
Career	.11	.05	.13	.15
Further study	.36**	.28**	.04	.07

* $p < .05$.
** $p < .01$.

TABLE 4

F TESTS FOR EVALUATION, INSTRUMENTALITY, AND PRODUCT SCORES

Items	Evaluation			Instrumentality			Product scores		
	Sex	Yr.	Sex × Y	Sex	Yr.	Sex × Y	Sex	Yr.	Sex × Y
Instrument	2.47			13.84**	4.02*		13.42**		
Importance	13.76**	3.87*		12.14**	2.69		14.11**		
Values	8.39**	6.81**		5.71*	2.46		10.16**		
Opposite sex	6.47*	2.49		33.61**			34.52**		
Campus group		10.61**			10.97**				
Class	13.17**				5.35*		6.05*		5.00*
Self-confidence		2.26		4.86*	4.04*		2.56		
Independence		4.96*		4.61*					
Success	23.98**		3.69	33.82**	2.67		11.89**	3.87*	6.64**
Information	4.43*			2.18					
Friends	3.30								
Limitations	2.50	2.56		3.60			4.06*		
Career	27.76**	2.27		31.57**	4.02*		34.07**		
Further study	14.78**			2.15	3.10		11.01**		

Note.—Fs of less than 2.00 are omitted.

* $F(1, 349) = 3.84, p < .05$.** $F(1, 349) = 6.64, p < .01$.

females. The only item significantly related to level of happiness among the females but not the males was "Contributing in a distinguished and meaningful manner to some campus group." Class differences were also evident: "Acquiring an appreciation of ideas," "Developing relationships with the opposite sex," "Finding a spouse," and "Having many good friends" were all related to happiness in both freshman groups but were not significantly correlated with level of happiness in either group of juniors. The correlations of level of happiness with the two components of the product score indicate that level of happiness is more closely associated with perceived instrumentality than it is with the rated importance of the goal; in most cases, the correlations with the latter score are less than .10.

Sex and Class Differences in Ratings

One of the primary considerations in developing this instrument was the need to differentiate among goals which motivate college students. Since males and females tend to go to college for different reasons (Douvan & Kaye, 1962), it was expected that males and females would differ in their ratings on many of the items. There was also some evidence from the Higgins study that the year in col-

lege would be an important source of variance. These data were analyzed using a two-factor analysis of variance design: the results of the F tests are presented in Table 4.² As expected, sex differences are prominent in all three sets of scores. With the exceptions of the items "Achieving academic distinction" and "Preparing for a career which requires further study beyond the BA or BS," all significant sex effects reflect higher ratings by the females. The year effects are less clear-cut; their failure to appear in the product scores is due to the fact that the evaluation scores decrease from freshman to junior year while the instrumentality scores tend to increase. In all instances of Sex × Year interaction, the scores of the males decrease from freshman to junior year, while those of the females increase. These results conform well with the data obtained by Higgins.

² Tables containing the means and standard deviations of the evaluation, instrumentality, and product scores on each of the 14 goal statements for each of the four sex/class groups have been deposited with the American Documentation Institute, as has a copy of the PICT. Order Document No. 9188 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

DISCUSSION

What are the implications of these results for the validity of the PICT and for the conception of the diversity of motivations among students which underlies its use? Although the correlations with the general measure of satisfaction were not high, they do indicate some area of overlap between the two measures. Since a student's estimate of satisfaction is probably based on the extent to which the University is meeting his primary goals, the instrumentalities associated with his five or six most salient goals might show a stronger relationship than does the total summed-product score.

The fact that students who indicate a higher average level of happiness perceive the University as more helpful in the process of goal attainment than do less happy students is encouraging. One would expect that students who are making satisfactory progress toward their goals would be happier. However, the work of Kendall (1954), Nowlis (1965), and Wessman and Ricks (1966) indicates that mood level may influence any evaluation of the kind required in the PICT. Therefore, it is probable that happier students would tend to see their goals as more effectively realized by what the University has to offer, simply by virtue of their general optimism. There is no direct evidence that increases in the perceived instrumentality of the University from freshman to junior year contribute to increases in average level of happiness during this period. However, the pattern of correlations of level of happiness with the individual goal products is consistent with what one would expect to find if the students were progressing toward goals which are important in fashioning an identity. The sex differences apparent in these data are consistent with what other studies have revealed as basic differences in male and female orientations to college. They are also consistent with other data obtained in the present study, most notably the fact that level of happiness was significantly related to grade-point average among the males but not among the females. The college years are typically the time when the personal identity is crystalizing, and the student is developing

a coherent set of personal goals and values. For the male, his life goals and his goals as a college student tend to be consistent; both are usually oriented primarily toward an occupational identity. The goals which are associated with level of happiness among the males only are clearly related to this occupational identity. For the female, however, her goals as a college student often are in conflict with her life goals. Even if she is heading for a professional career, gender role (with its attendant goals of wife and mother) is a more prominent component of her identity than it is for the male. The fact that the summed-product score did not significantly differentiate between happy and unhappy junior females may reflect the increasing diversity of life plans within this group as the time of decision about the future course of one's life draws near.

The PICT would seem to be adaptable to a variety of college and university environments. The statements of the goals are sufficiently general to apply to any college and are sufficiently diverse to reflect student motivations at most institutions. It would be an interesting measure to use in the study of various subgroups within the student culture. The perceptions and goals of these groups often differ radically from those of the population as a whole, and the PICT would offer one method for evaluating the extent of these differences.

REFERENCES

- BIRNEY, R. C., & TAYLOR, M. J. Scholastic behavior and orientation to college. *Journal of Educational Psychology*, 1959, 50, 266-274.
- CARLSON, E. R. Attitude change through modification of attitude structure. *Journal of Abnormal and Social Psychology*, 1956, 52, 256-261.
- DOUVAN, E., & KAYE, C. Motivational factors in college entrance. In N. Sanford (Ed.), *The American college*. New York: Wiley, 1962. Pp. 199-224.
- HIGGINS, G. C. A factor analytic study of some dimensions of the college experience. Unpublished doctoral dissertation, University of Rochester, 1963.
- KENDALL, P. *Conflict and mood*. Glencoe, Ill.: Free Press, 1954.
- MCCONNELL, T. R., & HEIST, P. The diverse college student population. In N. Sanford (Ed.), *The American college*. New York: Wiley, 1962. Pp. 225-252.

- NOWLIS, V. Research with the Mood Adjective Check List. In S. Tomkins & C. Izard (Eds.), *Affect, cognition, and personality*. New York: Springer, 1965. Pp. 352-389.
- PEAK, H. Attitude structure and motivation. In M. Jones (Ed.), *Nebraska symposium on motivation: 1955*. Lincoln: University of Nebraska Press, 1955. Pp. 149-189.
- ROSENBERG, M. J. Cognitive structure and attitudinal affect. *Journal of Abnormal and Social Psychology*, 1956, 53, 367-372.
- WESSMAN, A. E., & RICKS, D. F. *Mood and personality*. New York: Holt, Rinehart & Winston, 1966.

(Received April 14, 1966)

Manuscripts Accepted for Publication in the
Journal of Personality and Social Psychology

- Some Effects of Upward Mobile Status in Established and Ad Hoc Groups: Dorothy Gartner and Marvin A Iverson*: 154 Charles Street, East Williston, New York 11596.
- Retention of Affective Material: A Further Verification of the Intensity Hypothesis: Satrajit Dutta and Rabindra N. Kanungo*: Department of Psychology, Dalhousie University, Halifax, Nova Scotia, Canada.
- Longitude of Acceptance and Attitude Change: Empirical Evidence for a Reformulation: Alvin L. Atkins,* Kay Kujala, and James Bieri: Department of Psychology, Ferkauf Graduate School of Education, Yeshiva University, 55 Fifth Avenue, New York, New York 10003.
- Accuracy of Judgments of Personal Attributes from Speech: Raymond G. Hunt* and Tip Kan Lin: Department of Psychology, State University of New York, Buffalo, New York 14214.
- Increase in Psychological Anomie as a Result of Radical and Undesirable Change Expectancy: Gunars Reimanis*: Division of Social Sciences, Corning Community College, Corning, New York 14830.
- Will an Isolated Individual Advise Higher Risk Taking after Hearing a Discussion of the Decision Problem? Helmut Lamm*: Wirtschaftschochschule Mannheim, Sozialpsychologie, 68 Mannheim, den Germany.
- Vicarious Reinforcement and Noninstrumental Punishment in Observational Learning: Kenneth D. Craig*: Department of Psychology, University of British Columbia, Vancouver 8, Canada.
- Energizing Effects of Cognitive Dissonance upon Dominant and Subordinate Responses: Nickolas B. Cottrell* and Dennis L. Wack: Department of Psychology, Kent State University, Kent, Ohio 44240.
- Weapons as Aggression-Eliciting Stimuli: Leonard Berkowitz* and Anthony LaPage: Department of Psychology, Morrill Hall, Cornell University, Ithaca, New York.
- Recall and Pleasantness of Balanced and Unbalanced Cognitive Structure: Harold B. Gerard* and Linda Fleischer: Department of Psychology, University of California, Riverside, California 92502.

* Asterisk indicates author for whom address is supplied.

SOCIAL REINFORCEMENT, SOCIOECONOMIC STATUS, AND SUSCEPTIBILITY TO EXPERIMENTER INFLUENCE¹

MABEL L. SGAN

Wellesley College

The relative effects of 3 different kinds of social reinforcement on children's susceptibility to E influence on a preference task were explored in a between-subjects experimental design. 72 1st-grade children, representing the sexes and 2 socioeconomic levels (middle and working classes) in equal numbers, comprised the sample. Each of 2 female adults served with a random $\frac{1}{2}$ of the Ss under each of the experimental conditions. Susceptibility to E influence was indexed by the number of times S shifted an initial preference to agree with E's stated preference. The results confirmed the hypothesized effectiveness of nurturance and withdrawal of nurturance over unresponsive attention, but they failed to bear out the expectation that withdrawal of nurturance leads to greater susceptibility than consistent nurturance. Working class boys were significantly less susceptible to E influence, irrespective of treatment, than working-class girls and middle-class boys and girls.

The effects of adult nurturance and its subsequent withdrawal on children's behavior has long been a topic of interest in theoretical accounts of the socialization process and retrospective studies of parent-child relationships (e.g., Bronfenbrenner, 1960, 1961; Dollard & Miller, 1950; Mussen & Distler, 1959; Payne & Mussen, 1956; Sears, Maccoby, & Levin, 1957; Whiting & Child, 1953). Results from recent experimental investigations of this issue parallel in a provocative manner some of the findings and theoretical analyses reported earlier. Thus, a number of researchers have demonstrated that the experimental introduction of nurturance enhances both imitative and simple motor learning (Bandura & Huston, 1961; Bandura, Ross, & Ross, 1961; Rosenblith, 1961; Stevenson, 1965). While nurturance has been found to be a salient factor in the modification of children's behavior, observed relationships are typically qualified by sex of the subject as well as sex of the experimenter.

Less certain is the case for the withdrawal of nurturance or social isolation. Gewirtz

(1954) observed that children displayed more attention-seeking behavior in the presence of a nonresponsive adult than in the presence of an adult who devoted his complete attention to the child. In addition, the experimenters were found to be more effective with children of the opposite sex than with children of the same sex as themselves. Using female experimenters, Hartup (1958) found that among nursery school children, nurturance withdrawal stimulated faster learning than consistent nurturance for girls. Results for boys were not significant, although there were indications that highly dependent boys responded to the withdrawal of nurturance in the same fashion as did the girls. The findings supplied by Rosenblith (1959, 1961) on imitative learning in kindergarten children support partially Hartup's results, but are again complicated by sex differences between subject and experimenter. For boys only, withdrawal of attention by a male model was found to produce more imitation than consistent attention. Girls, on the other hand, displayed more imitation when they were exposed to consistent attention, regardless of sex of the model. An interesting study by Stein and Wright (1964) attests to the complexity of the effects of nurturance manipulations. Increases in imitation were observed when the child reacted to nurturance withdrawal with increased dependency or when it responded to continuous nurturance with de-

¹ This article is based on a dissertation submitted to the Graduate School of Cornell University in partial fulfillment of the requirements for the PhD degree. The author wishes to express her appreciation to the members of her dissertation committee for their generous counsel and encouragement: Urie Bronfenbrenner, chairman, Robert B. MacLeod, and Henry N. Ricciuti. The valuable assistance of Carol A. Millsom in the execution of the experiment is also gratefully acknowledged.

creased dependency. Although the findings on the effects of nurturance withdrawal thus far are ambiguous, the trends reported by these investigators are intriguing and invite further research on the issue.

The experiment reported here was aimed primarily at extending social reinforcement research to children's behavior in a contrived influence situation. With the notable exception of the work of Janis and his colleagues (Janis & Hovland, 1959), comparatively few studies have examined the developmental aspects of conformity and persuasibility in young children despite the vast amount of research conducted on adults. The lag is surprising, since the latter studies indicate that exploration of this fundamental social process among young children may prove to be a fruitful source of hypotheses concerning the important features of socialization, as well as of the phenomenon in question. Specifically, the present investigation sought to determine the relative effectiveness of three different kinds of social reinforcement in inducing shifts in children's preferences toward agreement with an adult: consistent nurturance, withdrawal of nurturance, and unresponsive attention. In line with the previous research cited, it was expected that, in general, children who had previously experienced consistent nurturance from a female adult on a simple preference task would subsequently be more susceptible to that adult's influence on a similar task than children who had experienced merely unresponsive attention. Further, it was predicted that withdrawal of nurturance would be more effective than either consistent nurturance or unresponsive attention in inducing susceptibility to experimenter influence. These expectations take as their point of departure the social learning formulations developed by the Yale group (e.g., Dollard & Miller, 1950; Mowrer, 1950; Sears, 1957). The view begins with the familiar argument that the nurturant responses of parents and other adults acquire generalized secondary reinforcing properties for the child through association with primary reinforcers (Dollard & Miller, 1950; Sears, 1957). Such affectional rewards are presumed to predispose the child to imitate the behavior of the adult model for the satisfaction

these cues provide (Mowrer, 1950). Once the parental characteristics have acquired positive reinforcing properties, the threat or the actual withdrawal of approval and support by an otherwise nurturant adult is believed to further stimulate the child to perform imitative responses. Thus, the child reinforces himself by continued support and a reduction in anxiety over the loss of nurturance. The child will respond with behavior that will, in his judgment, secure for him the desired warmth, approval, and support. He will tend to alter his behavior in an attempt to assure himself continued nurturance (Sears, 1957; Whiting and Child, 1953). Some empirical support for this prediction is available in the reported relationship between the use of discipline techniques involving the parent's withdrawal of love after the child's transgression and "high conscience" (Sears, Maccoby, & Levin, 1957). In summary, the present study predicted that the relative effectiveness of the three treatments in inducing susceptibility to experimenter influence would be as follows: nurturance withdrawal > nurturance > neutral.

Another variable which may contribute to differential behavior in the experimental situation is the socioeconomic background of the subject. Few studies have made explicit attempts to determine possible differences in the effect of social reinforcers as a function of social class membership. What little data exist on the issue are equivocal. Zigler and deLabry (Zigler, 1962) have observed that lower-class children were generally less responsive to verbal reinforcement in a marble-sorting task than were middle-class children. In a further analysis of two classes of verbal reinforcers, Zigler and Kanzer (1962) have reported that praise was less effective with middle-class children than were statements implying correctness of response, while the reverse relationship appeared with lower-class children. A replication of this study by Rosenhan and Greenwald (1965) failed to substantiate the major finding, although a significant interaction was found between sex of subject and socioeconomic status. Specifically, middle-class girls and lower-class boys were found to be more responsive to both kinds of reinforcers from a male experimenter

than were middle-class boys and lower-class girls. One explanation for the discrepant findings may lie in the different sampling procedures employed by the two sets of investigators. The Rosenhan-Greenwald results may possibly reflect interactions between the race and sex of examiners and subjects, inasmuch as Negro subjects were used for the lower-class group. In any event, the general social susceptibility of children of the two social classes remains an open issue. Surveys of child-rearing practices (Bronfenbrenner, 1958; Sears, Maccoby, & Levin, 1957) lead one to expect, however, that lower-class children would be less susceptible to influence attempts than their middle-class counterparts and, by inference, less responsive to nurturance and its withdrawal.

METHOD

Subjects

The subjects were 36 boys and 36 girls drawn from the entire first grade population of two public schools in Ithaca, New York. Two major subgroups were counterbalanced with respect to sex of subject and socioeconomic status (SES) as determined by a modification of Warner's scale of occupational status (Warner, Meeker, & Eells, 1949). The middle-class group was composed primarily of professional, business, and managerial occupations, together with other white-collar occupations such as salesmen and clerical workers. The working-class group included blue-collar workers, ranging from carpenters and plumbers to unskilled laborers; the largest number were semiskilled workers who were not self-employed. Children who were repeating the first grade were excluded from the sample. All subjects came from intact homes of white, native-born parents. Assignment to the three treatment conditions was random.

Experimenters

Each of two female graduate students majoring in child development served with one random half of the sex-SES subgroups under each of the experimental conditions. Although they were of similar age, one of the experimenters was Caucasian, the other was Oriental. Neither of the women had had previous contact with any of the subjects.

Measure of Susceptibility to Experimenter Influence

The susceptibility series was designed specifically to measure the dependent variable in this experiment. Drawn from a pool of approximately 100 pictures taken from current issues of American magazines, the 18 pairs of objects constituting the series were selected so as to be similar enough to preclude strong preference for one member of the pair

over the other, but not so similar as to destroy the reliability of preferences under noninfluence situations.

The items in the susceptibility series and the 14 pairs of a supplementary reinforcement series were selected from among the items which survived an item analysis of 56 pairs of pictures administered in several pretests. Subjects of the preliminary studies were the entire first-grade population of the public school systems of three towns in upstate New York. Items which showed significant sex differences were eliminated, as were those which were too similar to be discriminated reliably, as indicated by the children's verbalizations. All but three pictures each in the susceptibility and reinforcement series were in color.

Procedure

Employing a before-after design, the experiment consisted of three phases conducted in the following sequence: Phase I was administered in a morning session, Phases II and III in an afternoon session. The mean time interval between the two sessions was 3½ hours. Each subject was tested individually and participated once under only one of the experimental conditions. In each of the two schools from which the samples were drawn, the experimental room was a quiet, isolated office, customarily used by visiting specialists, including school psychologists, remedial reading instructors, and nurses.

Phase I consisted of a neutral administration of the susceptibility series. In an informal session aimed at providing a gamelike atmosphere, the subject was introduced to the experimental task with the following instructions:

I would like to play a game with you. We call the game: which do you like best? I will show you two pictures and you will tell me which you like the best. Then I will show you two other pictures, and you will tell me which you like the best. We will do this until the game is over. All right? Let's begin. Here are the first two pictures: which do you like the best?

During this session the experimenter played the role of an attentive but nonreinforcing observer. She displayed an interested but undemonstrative manner and recorded each of the child's preferences without comment. However, in order to set the subjects at ease in this situation, each child was told that he was "playing the game very nicely" after he had expressed his preference on the second item of the series. Except for this standard remark, no additional verbal support was given, and the experimenter took care to avoid nodding or saying "um-hmm." At the completion of the preference task, the subject was thanked for playing the game and sent back to his classroom. Subjects were asked to promise not to tell any other child about the game they played.

Phase II, administered after an interval of approximately 3½ hours, consisted of the experimental manipulation of reinforcement conditions immediately preceding attempts at influence. Subjects were

first introduced to the reinforcement series as follows:

Do you remember the game we played this morning? I have another game just like it. I will show you some pictures and you will tell me which you like the best and why you like it best. Okay? Let's begin. Here are the first two pictures: which do you like the best? Why do you like it best?

Then one of the three treatments was administered in the following manner:

1. Under the nurturance condition, the experimenter was rewarding, supportive, and encouraging throughout the treatment. She agreed with each of the child's preferences with verbal reinforcement and paraphrased in a supportive manner the reasons the child gave for his choices. In addition, she leaned toward the subject, looked directly at him, smiled, nodded, and made other approving comments at appropriate intervals. In each of the treatment conditions, attempts by the subject to engage the experimenter in conversation were pleasantly but quickly terminated.

2. Under the nurturance-withdrawal condition, the experimenter's behavior was similar to that displayed in the nurturance condition, but was limited to the first half of the treatment only; that is, during the administration of the first seven items of the reinforcement series. On the remaining items, the experimenter displayed nonsupportive behavior: she disagreed with each of the child's preferences and withheld all supportive paraphrases of the reasons offered by the child. In contrast to the nurturance condition, she leaned away from the subject, refrained from smiling and nodding, and avoided eye contact with the child.

3. Under the neutral condition, the experimenter's behavior was identical with that shown during the administration of the susceptibility series in Phase I. She provided no form of social reinforcement other than her attentiveness to the subject's preferences. She made no comments and avoided nodding or murmuring "um-hmm."

Although an attempt was made to keep constant the amount of time spent with each subject, the nature of the social interaction prescribed for each of the experimental conditions precluded complete achievement of this aim. In anticipation, therefore, of a possible significant regression of the dependent variable upon treatment time, a record was kept of the total length of time spent by each subject in Phase II.

Phase III, the critical influence situation, began immediately after the completion of the experimental manipulation of social reinforcement. The susceptibility series was readministered to all subjects and introduced with the following transitional remarks:

Now, this time, I will go first. I will tell you which picture I like best and then you tell me which one you like best. Okay?

The experimenter then presented each pair of pictures, saying, "I like this one [pointing to one member of each pair], which one do you like?" The

experimenter's stated preferences actually consisted of contradictions of 15 of the child's 18 preferences expressed in Phase I. In order to allay suspicion, the experimenter agreed with the subject's initial preferences on three dummy items. Throughout the session, the experimenter carefully avoided reinforcing the child's agreement or disagreement with the experimenter's preferences by withholding all forms of verbal and nonverbal reinforcement.

A procedural postscript. While it was the consensus of the experimenters that the nurturance-withdrawal condition did not evoke stressful reactions in the children, indirect cues provided by changes in voice quality, gross bodily movements, and hesitation before expressing preferences and verbalizing reasons suggested that the treatment probably did arouse a mild degree of anxiety or discomfort in some of the children. A repair measure was thus instituted at the end of the experiment in order to alleviate any residual anxiety in the subject which may have resulted from the previous experimental treatment. Accordingly, all subjects who had received withdrawal of nurturance were re-nurtured by the experimenter through a final administration of five pairs of items similar to those in the susceptibility and reinforcement series. The child was again instructed to indicate his preferences and to verbalize the bases of his choices; in each instance, the experimenter expressed agreement, made supportive comments, and interacted with the subject in a warm and friendly manner.

Measurement of the dependent variable. Susceptibility to experimenter influence was defined by the number of times the subject shifted an initial preference to agree with the experimenter's stated preferences. Since the responses on the three dummy items were excluded from consideration, it was possible for the scores to range from 0 (no shifts to agree with the experimenter's preferences) to 15 (shift of all initial preferences to agree with the experimenter's preferences).

RESULTS

Examination of the overall means and standard deviations for the separate sexes and for the combined sexes in the two schools revealed no differences approaching statistical significance. Consequently, the data from the two schools were combined for all subsequent analyses.

As another preliminary step, the relationship between the dependent variable and time spent in treatment (Phase II) was ascertained. Clearly, if time were correlated significantly with susceptibility scores, differences found for this variable could be spuriously determined by differences in the length of time spent in the experimental condition. Inspection of the relevant data indicated that although boys and girls did not spend signifi-

TABLE 1

ANALYSIS OF VARIANCE OF SUSCEPTIBILITY SCORES
ADJUSTED TO EXPERIMENTER, EXPERIMENTAL
CONDITION, SEX, AND SES

Source	df	MS	F
Experiment (E)	1	26.88	2.98
Experimental condition (EC)	2	72.85	8.01***
Sex of subject (S)	1	9.39	1.04
Socioeconomic status (SES)	1	29.38	3.25
E × EC	2	12.35	1.37
E × S	1	6.73	1.00
E × SES	1	29.40	3.26
EC × S	2	10.59	1.17
EC × SES	2	2.26	<1
S × SES	1	56.89	6.30*
E × EC × S	2	15.99	1.77
E × EC × SES	2	17.12	1.90
E × S × SES	1	.21	<1
EC × S × SES	2	27.21	3.01
E × EC × S × SES	2	13.24	1.47
Error (within)	48	9.03	
Total	71		

* $p < .05$.
*** $p < .01$.

cantly different amounts of time under each condition, there was, nevertheless, a tendency for time to increase from the neutral through the nurturance withdrawal to the nurturance conditions for both sexes. Product-moment correlations between susceptibility scores and treatment time for boys, however, did not differ significantly from chance ($r = .05$, $p > .05$); on the other hand, the obtained correlation for girls barely attained significance ($r = .32$, $p = .05$) thereby suggesting a tendency for those subjects who had spent more time in the treatment phase of the experiment to shift their preferences toward agreement with the experimenter more frequently than those who had spent less time.

Further analyses of the experimental data for girls by means of the covariance adjustment technique demonstrated that the differences in susceptibility scores for girls were significant beyond the .01 level even when scores were adjusted for differences in treatment time. The findings therefore permit one to conclude that any variations which may emerge between the experimental groups on the dependent variable cannot be attributed to differences between the subjects in the length of time spent in the reinforcement condition.

The difference in the ethnic backgrounds of the two experimenters prompted the decision to include the experimenter as an additional

dimension in the analysis of data, although it was not initially offered as one of the major experimental variables. Since it was conceivable that the differences in ethnicity could have operated to produce differential effects quite apart from the experimental manipulation of social reinforcement and social influence, and in the light of the literature on experimenter effects (e.g., Rosenthal, 1963, 1964; Stevenson, 1961), the separate data for the experimenters were evaluated independently by means of the analysis of variance.

Table 1 presents a summary of the analysis of the effects of experimenter, experimental condition, sex of subject, and SES level on susceptibility scores. The results indicated that the main effect of experimental condition was a highly significant source of variation. The F value of 8.01 ($df = 2$, $p < .01$) provides support for the inference that susceptibility to experimenter influence was affected by the preliminary social interaction. Table 2 shows that mean susceptibility scores tend to decrease from the nurturance to the neutral conditions. The more specific differences contributing to the significant F for the experimental condition were investigated by one-tailed t tests, whose error variances were based on the error term ($df = 48$) of the analysis of variance. The results summarized in Table 2 indicate that the mean of the neutral condition differs significantly from the means of both nurturance ($p < .0005$) and nurturance withdrawal ($p < .025$). The difference between the means for the latter conditions is significant at the .05 level, but contrary to prediction, nurturance was associated with higher susceptibility scores than

TABLE 2
COMPARISON OF MEAN SUSCEPTIBILITY SCORES
FOR EXPERIMENTAL CONDITIONS

Experimental condition	M	SD	t
Nurturance(N)	6.42	4.35	N vs. NW: 1.68*
Nurturance with- drawal(NW)	4.95	2.77	N vs. 0: 3.99****
Neutral(O)	2.96	2.51	NW vs. 0: 2.31**

Note.—Bartlett test of homogeneity of variance yielded a nonsignificant chi-square of 2.58, $p > .20$.

* $p = .05$, 1-tailed.

** $p < .025$, 1-tailed.

**** $p < .0005$, 1-tailed.

nurturance withdrawal. Thus, while the data support the hypothesis of the greater effectiveness of consistent nurturance from the experimenter as compared with unresponsive attention, they are less obliging with respect to the predicted superiority of withdrawal of nurturance in inducing subsequent susceptibility in these subjects. However, it is of interest to note the differential responses of the middle-class girls. For the neutral treatment, the mean susceptibility score was 2.33 ($SD = 1.49$); for nurturance and nurturance withdrawal, the means were virtually identical ($M = 6.00$, $SD = 2.58$; $M = 6.33$, $SD = 2.50$, respectively). This pattern of responses did not appear in any of the remaining sex-SES groups and suggests that the intent of the nurturance withdrawal treatment came closest to being realized among the middle-class girls.

The main effects for the experimenter and SES and the simple interaction between these two were of borderline significance ($p < .10$), reflecting a slight tendency for the overall susceptibility scores obtained by the Oriental experimenter (E_1) to be higher than those of the Caucasian experimenter (E_2). There was, moreover, a tendency for middle-class children to have somewhat higher scores than working-class children and, in addition, to be slightly more susceptible to E_1 than to E_2 .

Table 1 also shows a significant first-order interaction effect between sex and SES ($p < .05$). The finding is clarified by an inspection of the means shown in Table 3 and by the results of t tests which reveal that working-class boys have a significantly lower mean susceptibility score than working-class girls ($p < .05$), middle-class boys ($p < .02$), and middle-class girls ($p < .05$), irrespective of experimental condition; the means of the latter three groups do not differ significantly from one another. Inspection of the primary data indicated that the means for the three experimental conditions for working-class boys showed little variation ($M = 2.67$, $SD = 2.13$ for nurturance; $M = 3.50$, $SD = 2.81$ for nurturance withdrawal; $M = 2.50$, $SD = 2.63$ for neutral). Furthermore, there was no significant difference between the two experimenters in their effectiveness with working-class boys (for

TABLE 3
MEAN SUSCEPTIBILITY SCORES ACCORDING
TO SEX AND SES

SEX	Boys	Girls	Combined
Middle class	5.04 _a	4.89 _a	5.41
Working class	2.89 _b	3.89 _b	4.14
Combined SES	4.41	5.14	

Note.—Cells containing the same subscript are not significantly different at the .05 level by 2-tailed test.

E_1 , $M = 3.11$, $SD = 3.00$; for E_2 , $M = 2.67$, $SD = 2.05$). This finding eliminates the possibility that working class boys were reacting differentially to the Oriental experimenter because of greater prejudice.

Reliability of the Susceptibility Series

Evidence for the reliability of the susceptibility series was obtained by correlating each subject's susceptibility score on the odd items with his score on the even items. The corrected odd-even reliability coefficients for the nurturance, nurturance withdrawal, and neutral conditions were .78, .68, and .83, respectively; for the combined groups the overall reliability coefficient was .72. All of the reliability coefficients were significant at the .01 level. It is interesting to note that the reliability coefficient was lowest for the nurturance withdrawal condition. This finding may reflect the slightly unsettling effect of the previous social interaction resulting in mild discomfort and instability of choice behavior during the influence attempts.

DISCUSSION

That susceptibility to adult influence is enhanced by previous social reinforcement is clearly supported by these findings: children who had been exposed to consistent nurturance and rewarding interaction with the experimenter shifted their initial preferences to agree with that experimenter to a significantly greater degree than children who had been exposed to unresponsive attention. This finding is in accord with previous experimental demonstrations of the effectiveness of nurturance in the modification of children's behavior. It should be emphasized that the manipulation of this investigation (Phase II) did not involve reinforcement of imitative responses; what was reinforced was the state

of agreement and supportiveness between the adult and the child. The intent of the nurturance treatment was to achieve a certain degree of positive reinforcement for the adult-child interaction.

In demonstrating the effect of nurturance on susceptibility to experimenter influence, the present study also reveals the operation of the tendency for interpersonal behavior to be reciprocated. Thus, the degree of generosity evoked in subjects by another person's behavior is a function of the subject's perception of that person's own benevolent or malevolent intentions (Lambert, 1959). Dominative behavior on the part of one child induces domination in his partner; integrative behavior in one child elicits integrative behavior in his companion (Anderson, 1939). And in the present instance, agreement on the part of the adult induces agreement in the child.

It should be noted that while the explicit influence attempts of the experimenter in Phase III involved neither coercion nor reinforcement for agreement, rewards were not entirely absent from the influence situation. The experimenter with whom the child could agree or disagree was not only prestigious by virtue of age and size, but also her identification with teachers and other school authorities was emphasized by the fact that the experiment was being conducted within the school setting, where compliance to adult figures typically receives positive reinforcement. Yet the significantly lower degree of susceptible behavior displayed by the children in the neutral condition suggests that the influence of these additional reinforcers was minimal and points instead to the differential social reinforcement manipulated by the experimenter as the critical variable.

The failure of the experimental results to support the predicted relation between consistent nurturance and the withdrawal of nurturance requires further comment, in view of the positive results reported by previous investigators. Underlying the prediction was the assumption that an implied threat to the relationship between the child and the experimenter following a period of nurturance would arouse a certain amount of anxiety in the child which, in turn, would motivate shifting

to agree with the experimenter. The finding that nurturance and nurturance withdrawal were equally effective with middle class girls is suggestive in light of previous studies which show that it is these girls, rather than working-class children or middle class boys, who are most likely to be exposed to "love-withdrawal" discipline techniques (Bronfenbrenner, 1961; Sears, Maccoby, & Levin, 1957).

One explanation for the failure of the nurturance-withdrawal hypothesis to find confirmation may lie in the form of threat employed. It is quite possible that disagreement with 7 out of 14 choices and the withholding of supportive paraphrases, comments, and smiles by the experimenter had only mildly threatening implications for these children. Moreover, the fact that the experimenter remained in close view may have mitigated the treatment and may even have constituted a condition of mild support. The subtle nature of the technique used to imply withdrawal of nurturance in this experiment becomes apparent when it is examined against the methods employed in previous investigations of social deprivation and inconsistent nurturance on children's behavior. The latter techniques are characterized by a sharp termination of social reinforcement, including physical absence of the reinforcing figure for a specified period of time. Thus, the relatively mild and superficial nature of the treatment may have vitiated the experimental manipulation of nurturance withdrawal.

On the other hand, the work of Crandall and her associates (Crandall, 1963; Crandall, Good, & Crandall, 1964) on the sequence effects of social reinforcement raises the possibility that the treatment was not necessarily as mild or as superficial as has been suggested. Crandall found that the experimenter's silence following positive reinforcement had the effect of lowering the expectancy of success in an ambiguous situation, whereas negative reactions followed by silence led to the raising of expectancy. She posits a contrasting process whereby the child interprets an adult's silence as containing information opposite to that of the preceding reaction.

More plausible explanations for the dis-

repant results obtained in this study may be found in other dissimilarities in methodology. The difference in the chronological age of the subjects—preschoolers (Hartup, 1958; Stein & Wright, 1964) and kindergarteners (Rosenblith, 1961) as opposed to first graders in this study—may be relevant, if one considers the possibility that the withdrawal of motherly nurturance has less threatening implications for 6- to 7-year-old children than for 3- to 4-year-old subjects. Finally, there was a significant difference in the criterion task used in this experiment. Stein and Wright (1964) and Rosenblith (1961) used simple imitation, whereas in this study, an imitative response demanded that the child change his original preference. In addition, the criterion tasks of previous research were designed to elicit adult approval; in this experiment, social reinforcement was explicitly withheld from the criterion task.

The predicted difference between the two social classes was confirmed by the behavior of the working-class boys only. As indicated earlier, these subjects showed little variation in their responses to the three experimental conditions. This fact, coupled with their significantly lower mean susceptibility score in comparison with the other three sex-SES subgroups, suggests the hypothesis that the working-class boys may have been actively resisting the experimental manipulation of both social reinforcement and social influence. The present study would have been strengthened methodologically by the inclusion of a control group in which children were administered the susceptibility series twice without any intervening reinforcement condition or influence manipulation. Such a control would not only have provided data on the test-retest reliability of the series, but would also have furnished a base level against which the experimental effects could be assessed more systematically. Had such a group been included, it would be expected that the change score of the working-class boys would be significantly less than that of the unrun control group, if they were, indeed, consciously resisting experimenter influence.

This experiment was not originally designed to assess the relative adequacy of the social drive (Berkowitz, 1964; Erickson,

1962; Gewirtz & Baer, 1958a, 1958b) as opposed to the valence position (Berkowitz, Butterfield, & Zigler, 1965; Berkowitz & Zigler, 1965; McCoy & Zigler, 1965; Shallenberger & Zigler, 1961) in accounting for the differential effectiveness of social reinforcers following different types of social interaction. Nevertheless, the present findings do appear to lend some post hoc support to the valence position. Within the social drive framework, interaction with a highly nurturant adult is viewed as satiating social drive and leading to a decrease in the child's responsiveness to the social reinforcers which the adult dispenses, whereas deprivation of social reinforcers is seen as increasing social drive and enhancing the adult's effectiveness as a reinforcing agent. According to the valence interpretation, the child's responsiveness to the social reinforcers which the experimenter dispenses is a direct function of the valence which the experimenter has for the child. This valence, or the child's attitude toward the experimenter, is, in turn, determined by the nature of the social contact. A positive interaction is thus viewed as increasing the experimenter's effectiveness and a negative interaction as reducing it. In line with the valence position, the present study suggests that it was the magnitude of the dose of nurturance the child received rather than the sequence of nurturance followed by its withdrawal which enhanced the experimenter's effectiveness as an influencing agent.

REFERENCES

- ANDERSON, H. H. Domination and social integration in the behavior of kindergarten children and teachers. *Genetic Psychological Monographs*, 1939, 21, 287-385.
- BANDURA, A., & HUSTON, A. C. Identification as a process of incidental learning. *Journal of Abnormal and Social Psychology*, 1961, 63, 311-318.
- BANDURA, A., ROSS, D., & ROSS, S. A. Transmission of aggression through imitation of aggressive models. *Journal of Abnormal and Social Psychology*, 1961, 63, 575-582.
- BERKOWITZ, H. Effects of prior experimenter-subject relationships on reinforced reaction time of schizophrenics and normals. *Journal of Abnormal and Social Psychology*, 1964, 69, 522-530.
- BERKOWITZ, H., BUTTERFIELD, E. C., & ZIGLER, E. The effectiveness of social reinforcers on persistence and learning tasks following positive and negative interactions. *Journal of Personality and Social Psychology*, 1965, 2, 706-714.

- BERG, WITZ, H., & ZUCKER, E. Effects of preliminary positive and negative reinforcement and delay reinforcement on children's responses to social reinforcement. *Journal of Personality and Social Psychology*, 1967, 2, 10-15.
- BRINTENBERGER, U. Social class and social class structure. In L. P. Lipsitt, E. E. Maccoby, I. M. Newman, & E. L. Hartley (Eds.), *Research in social psychology*. 2nd ed. New York: Holt, Rinehart & Winston, 1958. Pp. 400-425.
- BRINTENBERGER, U. Frustration theories of identification and their derivatives. *Child Development*, 1960, 31, 15-40.
- BRINTENBERGER, U. Same factor? Interrelatedness of responsibility and leadership in adolescents. In L. P. Lipsitt & B. M. Bass (Eds.), *Leadership and interpersonal behavior*. New York: Holt, Rinehart & Winston, 1961. Pp. 269-272.
- CRANDALL, V. C. Reinforcement effects of adult reactions and motivations on children's achievement expectations. *Child Development*, 1960, 34, 388-384.
- CRANDALL, V. C., GOOD, S., & CRANDALL, V. J. Reinforcement effects of adult reactions and motivations on children's achievement expectations; a replication study. *Child Development*, 1964, 35, 485-497.
- DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
- ERICKSON, M. T. Effects of social deprivation and satiation on verbal conditioning in children. *Journal of Comparative and Physiological Psychology*, 1962, 55, 955-957.
- GEWIRTZ, J. L. Three determinants of attention-seeking in young children. *Monographs of the Society for Research in Child Development*, 1954, 19(2, Whole No. 59).
- GEWIRTZ, J. L., & BAER, D. M. Deprivation and satiation of social reinforcers as drive conditions. *Journal of Abnormal and Social Psychology*, 1958, 57, 165-172. (a)
- GEWIRTZ, J. L., & BAER, D. M. The effect of brief social deprivation on behavior for a social reinforcer. *Journal of Abnormal and Social Psychology*, 1958, 56, 49-56. (b)
- HARTUP, W. W. Nurture and nurture-withdrawal in relation to the dependency behavior of preschool children. *Child Development*, 1958, 29, 191-201.
- JANIS, I. L., & HOVLAND, C. I. (Eds.) *Personality and persuasibility*. New Haven: Yale University Press, 1959.
- LAMBERT, N. W. An experimental study of benevolence and malevolence. Unpublished doctoral dissertation, Cornell University, 1959.
- MCCOY, N., & ZIGLER, E. Social reinforcer effectiveness as a function of the relationship between child and adult. *Journal of Personality and Social Psychology*, 1965, 1, 604-612.
- MOWLER, O. H. Identification: A link between learning theory and psychotherapy. In *Learning theory and personality dynamics*. New York: Ronald Press, 1950. Pp. 573-616.
- MUSSEN, P., & DISTLER, L. Masculinity, femininity, and identification. *Steps of Journal of Normal and Social Psychology*, 1959, 59, 350-355.
- PAYNE, D., & MUSSEN, P. Parent-child relations and father identification. *Journal of Abnormal and Social Psychology*, 1956, 52, 358-367.
- ROSENBLITH, J. F. Learning by imitation in kindergarten children. *Child Development*, 1959, 30, 69-80.
- ROSENBLITH, J. F. Imitative color choices in kindergarten children. *Child Development*, 1961, 32, 211-223.
- ROSENTHAL, D., & GREENWALD, J. A. The consequences of sex and socioeconomic class on the effectiveness of two classes of verbal reinforcement. *Journal of Personality*, 1965, 33, 108-121.
- ROSENTHAL, R. Experimenter attributes as determinants of subjects' responses. *Journal of Personality Techniques and Personality Assessment*, 1963, 27, 324-331.
- ROSENTHAL, R. The effect of the experimenter on the results of psychological research. In B. A. Maher (Ed.), *Progress in experimental personality research*. New York: Academic Press, 1964. Pp. 79-114.
- SEARS, R. R. Identification as a form of behavioral development. In D. B. Harris (Ed.), *The concept of development*. Minneapolis: University of Minnesota Press, 1957. Pp. 149-161.
- SEARS, R. R., MACCOBY, E. E., & LEVIN, H. *Patterns of child rearing*. Evanston, Ill.: Row, Peterson, 1957.
- SHALLENBERGER, P., & ZIGLER, E. Rigidity, negative response tendencies, and co-satiation in normal and feeble-minded children. *Journal of Abnormal and Social Psychology*, 1961, 63, 20-26.
- STEIN, A. H., & WRIGHT, J. C. Imitative learning under conditions of nurturance and nurturance withdrawal. *Child Development*, 1964, 35, 927-938.
- STEVENSON, H. W. Social reinforcement with children as a function of CA, sex of E, and sex of S. *Journal of Abnormal and Social Psychology*, 1961, 63, 147-154.
- STEVENSON, H. W. Social reinforcement of children's behavior. In L. P. Lipsitt & C. C. Spiker (Eds.), *Advances in child development and behavior*. Vol. 2. New York: Academic Press, 1965. Pp. 97-126.
- WARNER, W. L., MEKKER, M., & EELIS, K. *Social class in America*. Chicago: Science Research Associates, 1949.
- WHITING, J. W. M., & CHILD, I. L. *Child training and personality: A cross-cultural study*. New Haven: Yale University Press, 1953.
- ZIGLER, E. Rigidity in the feeble-minded. In E. P. Trapp & R. Himelstein (Eds.), *Research readings on the exceptional child*. New York: Appleton-Century-Crofts, 1962.
- ZIGLER, E., & KANZER, P. The effectiveness of two classes of verbal reinforcers on the performance of middle- and lower-class children. *Journal of Personality*, 1962, 30, 157-163.

(Received April 27, 1966)

EFFECTS OF FACILITATING TECHNIQUES AND SUBJECT-EXPERIMENTER INTERACTION ON CREATIVE PROBLEM SOLVING

MEREDITH GALL AND GERALD A. MENDELSON¹

University of California, Berkeley

Effects of 3 procedures for facilitating creative problem solving after an initial impasse has been reached were compared. In the 1st procedure incubation, Ss' attention was diverted for a time from the problem. Drawn from Mednick's, 1962, Remote Associates Tests by means of a converbal task, in the 2nd, they continued to work on the problems, and in the 3rd, Ss free associated to the elements of the problem. In 2 experiments, continued work produced significantly more solutions than incubation for all Ss. Association training had a facilitating effect for females only. Their performance was, in addition, strongly influenced by the sex of E. The results raise questions about the efficacy of incubation in facilitating convergent solutions and point to the importance of social psychological variables as determinants of performance, at least for females.

Despite the upsurge of interest in the study of creativity since 1950, there has been relatively little empirical work directed toward the problem of finding techniques to facilitate creative performance. Maltzman's conclusion in 1960 that "most of the work in this important area has been anecdotal or merely hortatory [p. 229]" seems equally appropriate today. This is not to say that psychologists and other writers have ignored this problem, but rather that the amount of relevant objective information obtained under controlled conditions is small.

Historically, the most frequently noted facilitating procedure is an incubation period following initial work on a problem or piece of art. Although the importance of incubation seems generally accepted, and numerous explanations of how it operates have been proposed (Mednick, Mednick, & Mednick, 1964; Poincaré, 1952; Woodworth & Schlosberg, 1954; Youtz, 1962), Crutchfield (1961) notes the "embarrassing fact" that "... we do not really know with any assurance whether there is *any* beneficial effect of incubation on creative thinking [p. VI-11]." He goes on to point out that continued work on a problem may be as effective or more effective

in producing a solution than a period away from the problem. One of the purposes of the present research is to evaluate this possibility.

The work of Maltzman (Maltzman, Beloni, & Fishbein, 1964; Maltzman, Bogartz, & Breger, 1958; Maltzman, Brooks, Bogartz, & Summers, 1958; Maltzman, Simon, Raskin, & Licht, 1960) provides the most thoroughly investigated technique for facilitating originality.² The essence of his procedure is to train subjects to produce original (statistically infrequent) free-associative responses by requiring them to give a different association to each repeated presentation of the same stimulus word. There is little doubt that this procedure does produce more original associations during the course of the training period (Caron, Unger, & Parloff, 1963; Maltzman, Bogartz, & Breger, 1958), but the extent to which the originality generalizes to new situations seems to be dependent upon the similarity of the transfer task to the training procedure. Thus, the most pronounced and consistent effects of training are on free associations to a different list of stimulus words presented subsequent to the training period. Similar but less conclusive

¹ The authors wish to express their thanks to Donald W. MacKinnon and the staff of the Institute of Personality Assessment and Research, where this paper was prepared and written, for their support and helpful criticism.

² Maltzman is careful to distinguish between originality and creativity, but he, like other authors (cf. Guilford, 1959), sees originality as correlated with creativity; that is, it is necessary but not a sufficient condition for creativity.

results are obtained when Guilford's Unusual Uses Test is the transfer task (Maltzman, Bogartz, & Breger, 1958, Maltzman et al., 1960). It should be noted that both free association and the Unusual Uses Test call for divergent responses. In contrast, originality training has no apparent facilitating effect on the solution of problems requiring a convergent response—the two-string problem (Maltzman et al., 1964) or Mednick's (1962) Remote Associates Test (RAT—Caron et al., 1963; Maltzman et al., 1964). The failure of originality training to increase solutions on the RAT is of particular interest because successful performance on the test presumably entails the production of remote (statistically infrequent) associations. However, Freedman (1965) was able to facilitate RAT performance by using a continuous free-association training procedure; that is, unlike Maltzman's standard procedure in which the subject gives a discrete response to each stimulus presentation, here the subject was required to continue free associating for 30 seconds to each of 10 stimulus words. Since RAT solution likewise entails continuous rather than discrete association, the discrepancy between Freedman's results and those of the previous studies would seem to be attributable to the greater similarity of the nature of his training procedure to the nature of the transfer task.

The purpose of the experiment to be reported was to compare the effectiveness of two techniques of facilitating creative problem solving, incubation and free-association training, to each other and to a third condition, continued work on the problem. RAT solution was used as the dependent measure both to provide continuity with previous work and because of its presumed relevance to creativity.

METHOD

Two experiments were conducted, the second being an attempt to replicate the first. The procedure in the two studies was identical with the exception that different experimenters were used in the two studies.

Subjects and Experimenters

In each experiment, the subjects were 30 male and 30 female students, drawn from several undergraduate psychology courses for which participation

in an experiment was a requirement. Four experimenters, two male and two female, were used in each experiment; all were relatively advanced psychology majors who were paid for their work but had no knowledge of the purpose of the experiment. Since our pretest data suggested the possibility of sex of experimenter-sex of subject interactions, as did Freedman's (1965), male and female experimenters ran equal numbers of male and female subjects. One male and one female experimenter ran 7 male and 8 female subjects apiece, while the other pair of experimenters ran 8 male and 7 female subjects apiece. Of the 15 subjects run by each experimenter, 5 were in an incubation condition, 5 in a continued-work condition, and 5 in an association-training condition.

Remote Associates Test

Mednick (1962) provides a detailed description of the RAT and the theory of creativity it is designed to operationalize. The test seems particularly apropos for studies of free-association originality training because it demands the production of remote associates in the process of solution. The subject's task in the test is to find an associate common to three stimulus words; for example, "slow" is the solution to the item, GO POKE MOLASSES. It should be noted that the test involves both forward and backward associates, and thus it is difficult to assess the remoteness of the associations required for solution with any precision. In the standard administration, 30 items to be solved in 40 minutes are presented in a booklet, but in the present experiment an altered technique of presentation, described below, was used.

Procedure

Each subject was run individually in a session lasting approximately 2 hours. All subjects, regardless of the experimental condition to which they were assigned, were first given the RAT in the following manner. The 30 items of the test were printed separately on 3 × 5-inch cards. They were then presented to the subject in the order in which they appear in the booklet with these instructions: "You are to work on each problem until you solve it or two minutes have passed. Feel free to guess; I will tell you if your answer is correct." The experimenter wrote down whether the subject solved the problem and the time taken for solution. When the test had been completed, the experimenter set aside the first five problems that the subject had failed to solve. Immediately afterwards, the subject was run under one of three experimental conditions.

Condition 1 (incubation). After completing the test, subjects in this condition were asked to make psychophysical judgments about weights for 25 minutes. This task was used because it is non-verbal, and hence not likely to cause verbal interference or to provide unintended cues to problem solution, but demands sufficient attention that subjects could not easily think about the RAT problems on which they had been working. After the 25

minutes had elapsed, the subject was told, "I want you to work again on some of the problems that you previously missed. As before, you will have two minutes for each problem, and you can feel free to guess." The five problems were then administered, and the experimenter recorded whether or not they were now solved.

This procedure conforms to the usual description of incubation in that the subject's attention is diverted for a time from a task on which he has reached an impasse, and then subsequently he resumes work on the problem. Since there is no information on how long an incubation period should be, the choice of 25 minutes is essentially an arbitrary one, dictated by the necessity of completing the experiment in 2 hours.

Condition II (continued work). In this condition, immediately after the RAT had been completed, the subject was given the first five problems that he had failed to solve and was told to continue working on them. They were presented one by one with a time limit of 5 minutes for each. Thus, the subject had a possible 25 minutes to work on the problems, a time period comparable to that subjects in Condition I spent incubating. Of course, if the subject solved some of these problems during this period, the full 25 minutes was not used. After the subject had either solved or worked for 5 minutes on each problem, he was asked to work on those of the five that he still had not solved for an additional 2 minutes apiece. The purpose of this last 2-minutes-per-problem stage was to make all conditions as comparable as possible in total time. Thus, in this condition, the subject first had a 25-minute work period and then a final 10-minute work period, although both periods were diminished in time if the subject solved any of the problems.

Condition III (association training). In this condition, immediately after the RAT had been completed, subjects were asked to associate (for a total of 25 minutes) to the words comprising the five problems they had missed. Each card was presented separately to the subject and he was asked to give an associate to each of the three words regardless of whether he thought the associate was the answer word or not. Each of the five cards was presented in this manner, and then the sequence repeated. The subject's associations were recorded to insure that he did not give the same association twice; if the subject repeated an association, he was asked to give another in its place. The subject was also told that while he was associating he might come upon the answer word, but that if he did so, the experimenter would not tell him it was the correct answer; the subject himself had to recognize that it was the answer. During the 25 minutes, the typical subject went through the five cards six or seven times. After the 25 minutes were up, the subject was asked to work on the problems he had not yet solved, again for a period of 2 minutes each. As before, the purpose of this last stage was to make all three conditions as nearly comparable in time as possible.

The association-training procedure used in this

experiment more closely resembles that of Maltzman than that of Freedman, but with the important difference that the stimulus words are the same as the words which comprise the RAT problem. In this respect, the training and transfer tasks are quite similar. Another important difference between this study and previous studies is that the facilitation procedure follows rather than precedes the initial problem-solving period. This sequence was determined by our interest in incubation which, by definition, is a process occurring after an impasse has been reached. It should be noted, however, that with respect to the incubation procedure, the experiment was designed to provide an answer to the question of whether or not continued work on a problem is as effective or more effective in producing a solution than an equal period away from the problem. A different design would have been necessary to answer the question of whether "... there is any beneficial effect of incubation on creative thinking."

RESULTS

The data of the first experiment and the replication experiment were analyzed by means of a four-way analysis of variance in which the main effects were experiment, facilitating condition, sex of experimenter, and sex of subject. The means for both experiments are shown in Table 1 and the results of the analysis in Table 2.

With respect to the central question, the comparison of facilitation techniques, it is clear that the continued-work condition produces the greatest number of solutions and the incubation condition the fewest. This re-

TABLE 1
MEAN NUMBER OF SOLUTIONS IN ORIGINAL
AND REPLICATION EXPERIMENTS*

E	S	Experiment I			Replication experiment		
		Condition			Condition		
		I	II	III	I	II	III
Male	Male *	2.00	3.20	2.20	2.00	3.40	1.40
	Female	2.20	3.00	1.00	2.40	3.20	3.80
Female	Male	.80	3.00	1.80	1.80	2.20	1.40
	Female	1.20	2.40	4.00	1.00	2.00	1.60
	M	1.55	2.90	2.25	1.80	2.70	2.05

* Since the number of subjects in each cell is the same ($N = 5$), the reader can obtain the means for any main or interaction effect by calculating the mean of the appropriate means. Condition I is incubation, II, continued work, and III, association training.

TABLE 2
ANALYSIS OF VARIANCE

Source	df	MS	F
Experiment (A)	1	.08	<1
Conditions (B)	2	12.76	11.22****
Sex of E (C)	1	9.08	7.97****
Sex of S (D)	1	1.41	1.24
A × B	2	.68	<1
A × C	1	7.01	6.16***
A × D	1	.21	<1
B × C	2	3.23	2.83*
B × D	2	3.81	3.35**
C × D	1	.01	<1
A × B × C	2	4.61	4.05***
A × B × D	2	1.06	<1
A × C × D	1	9.08	7.97****
B × C × D	2	.81	<1
A × B × C × D	2	5.92	5.21****
Error	96	1.14	

* $p < .10$.
 ** $p < .05$.
 *** $p < .025$.
 **** $p < .01$.
 ***** $p < .001$.

sult is consistent across the two experiments; the Conditions × Experiment interaction is negligible. Given the highly significant F ratio for the conditions effect, the Newman-Keuls procedure for testing differences between ordered means (Winer, 1962) was applied to these data. The results indicate that the continued-work condition is significantly different ($p < .01$) from both the association-training and the incubation conditions which are, in turn, significantly different from each other ($p < .05$). It should be noted, however, that the Conditions × Sex of Subject interaction is also significant. Inspection of Table 1 indicates that for males the continued-work condition is superior to both the incubation and association-training conditions which are not different from each other, while for females the mean number of solutions in the continued-work and association-training conditions are equal to each other and greater than the mean for the incubation condition. The obtained sex difference is attributable, then, to the association-training condition which produces significantly ($p < .01$) more solutions for females than for males. The insignificant Experiment × Conditions × Sex of Subject interaction indicates that again the result is consistent across the two experiments.

The remaining significant main or interaction effects all involve sex of experimenter. First, male experimenters, in general, elicit

more solutions than do female experimenters, irrespective of the sex of the subject. This result, however, is far more pronounced in the replication than in the initial experiment. If we turn next to a consideration of sex matching, the situation becomes more complex. Male subjects solve more problems when paired with male experimenters in both experiments, but for females the results of the two experiments are inconsistent. In the first experiment, female subjects do better with female experimenters, but in the replication their performance is far superior when the experimenter is a male. It is evident that this result is attributable to the female subjects in the association-training condition: in both experiments, the performance of females in the incubation and continued-work conditions is superior when the experimenter is a male, but in the association-training condition, the means for female subjects in Experiment I are 1.00 and 4.00 for male and female experimenters, respectively, in contrast to the means of 3.80 and 1.60 obtained in the replication experiment.

In order to pursue these results in greater detail, separate analyses of variance were done for male and female subjects. The results are shown in Table 3. It is clear from these analyses that the subject-experimenter matching is a considerably more powerful influence on the performance of females than of males. For the latter, only the conditions main effect is clearly significant, while for the former all main and interaction effects involving sex of experimenter reach significance. Perhaps the best way to summarize these

TABLE 3
ANALYSES OF VARIANCE FOR EACH SEX

Source	df	Males		Females	
		MS	F	MS	F
Experiment (A)	1	.27	<1	.01	<1
Condition (B)	2	10.85	8.90****	5.72	5.40***
Sex of E (C)	1	4.27	3.50**	4.81	4.54**
A × B	2	1.62	1.33	.24	<1
A × C	1	.06	<1	16.03	15.12****
B × C	2	.42	<1	3.62	3.42**
A × B × C	2	1.32	1.08	9.21	8.69****
Error	48	1.22		1.06	

* $p < .10$.
 ** $p < .05$.
 *** $p < .01$.
 **** $p < .005$.
 ***** $p < .001$.

findings is as follows: for male subjects, 36% of the total sums of squares is explained, of which only 24% is accounted for by the main and interaction effects in which sex of experimenter figures; for females, 53% of the total sums of squares is explained, of which 80% is accounted for by the sex of experimenter main and interaction effects. Clearly, the social factors in the experimental situation are of much more importance to female subjects; the males, in contrast, appear more task oriented.

Finally, the hypothesis that the obtained results are attributable in some way to initial differences among subjects in RAT solving ability was tested by a four-way analysis of covariance in which RAT score was used as the covariate. The results do not differ to an appreciable extent from those shown in Table 2. In no case does a significant F ratio obtained in the analysis of variance become insignificant in the analysis of covariance, and moreover the within-cells regression of solutions on RAT scores is insignificant. Further, there is no evidence that subjects in any particular cell of the experimental design were given more or less difficult items to work on than subjects in any other cell.

DISCUSSION

Of the several statistically significant findings in these studies, three seem of particular importance. First, for both sexes, a period of continued work on a problem produces more solutions than an equal period of incubation. Second, the association-training procedure is effective for female subjects but not for male subjects. Third, the social psychological aspects of the experiment have a marked effect on the performance of the female subjects but a negligible one on the performance of the male subjects. We will discuss each in turn.

As noted in the introduction, Crutchfield (1961) points out that (a) we are not certain that incubation has a facilitative effect on creative thinking; and (b) a period of continued work on a problem may be as effective in producing a solution as an equal period of incubation. We pointed out before that the present study does not allow an assessment of the absolute effect of incubation—it is im-

possible to judge whether the fact that subjects do solve some of the problems after the incubation period is attributable to the time away from the problem or to the continued work on the problem during the 10-minute period following incubation. There is little doubt, however, about the second point; continued work is not merely as effective as incubation in producing solutions, it is markedly superior to incubation in producing solutions. There is naturally a question of whether incubation as operationalized here conforms to the meaning intended by those who write of its importance. This is a difficult question to answer, but as Woodworth (1938) argues:

The word incubation rather implies the theory of unconscious work on a problem during a period of attention to other matters, but we can strip off this implication and use it simply to denote the fact—so far as it is a fact—that a period of inattention to a problem intervenes after preparation and before illumination [p. 819].

In this sense, at least, the experiment is concerned with incubation. However, the RAT requires convergent solutions; the question of whether or not continued work would be equally superior to incubation in producing divergent solutions can be answered only by future investigations.

The data are not sufficient to allow for anything but speculation about the reasons for the greater effectiveness of the association-training condition for females. Previous work on sex differences in problem-solving behavior, summarized, for example, by Tyler (1965), suggests a tentative explanation, however. Females are typically less successful than males in solving problems which demand active restructuring of stimulus elements; moreover, females who score in the direction of masculinity on a personality inventory are superior on such problems to females scoring in the direction of femininity (Milton, 1957). The association-training procedure provides the subjects with a technique which demands the active restructuring of the elements of the problem, that is, the RAT item words and their relevant associations. For females, this may represent a new and potentially useful problem-solving strategy, while for males it may be little different from their habitual mode of approach. It is interesting in this

regard that in the one experiment³ in which Maltzman and his associates (Maltzman, Brooks, Bogartz, & Summers, 1958) were able to facilitate solutions of the two-string problem, the effect was substantially more pronounced for female than for male subjects. Clearly, more data bearing specifically on this line of reasoning are needed before conclusions can be reached.

The greater effect of the social psychological aspects of the experiment on female subjects than on males is to be expected in light of the extremely consistent finding that females are more dependent upon, sensitive to, and responsive to other people than are males (Tyler, 1965). It is particularly significant that in the present experiment, the response of females was most variable in the association-training condition, the condition in which there was maximum subject-experimenter social interaction. In his study of association training, Freedman (1965) after finding evidence of the impact of this interaction in his pretest data, decided to have all men run by male experimenters and all women by female experimenters. The present data suggest that the situation is more complex than that, however, for it appears that although sex matching is a powerful influence on the performance of females, its effects can just as easily be facilitating as inhibiting. We cannot specify what factors are involved, but the methodological significance of the finding is unmistakable. If, as in the usual case, the sex of experimenters had not been systematically varied and no replication attempted, conclusions about the effectiveness of the association-training condition could have been grossly incorrect.

It should be noted, in conclusion, that it would be as incorrect to exaggerate as to ignore the social psychological aspects of this experiment. First, they seem to have little effect on the males, and second, certain results emerge with strength, clarity, and consistency despite the subject-experimenter interactions. Specifically, continued work is superior to incubation in producing solutions to problems, and the association-training procedure, which in this experiment is quite simi-

lar to the demands of the transfer task, is effective, but only for females.

REFERENCES

- CARON, A. J., UNGER, S. M., & PARLOFF, M. B. A test of Maltzman's theory of originality training. *Journal of Verbal Learning and Verbal Behavior*, 1963, 1, 436-442.
- CRUTCHFIELD, R. S. The creative process. In *Conference on "The creative person."* Berkeley: University Extension, University of California, Liberal Arts Department, 1961. Pp. VI, 1-16.
- FREEDMAN, J. L. Increasing creativity by free-association training. *Journal of Experimental Psychology*, 1965, 69, 89-91.
- GUILFORD, J. P. Traits of creativity. In H. H. Anderson (Ed.), *Creativity and its cultivation*. New York: Harper, 1959. Pp. 142-161.
- MALTZMAN, I. On the training of originality. *Psychological Review*, 1960, 67, 229-242.
- MALTZMAN, I., BELLONI, M., & FISHBEIN, M. Experimental studies of associative variables in originality. *Psychological Monographs*, 1964, 78(3, Whole No. 580).
- MALTZMAN, I., BOGARTZ, W., & BREGER, L. A procedure for increasing word association originality and its transfer effects. *Journal of Experimental Psychology*, 1958, 56, 392-397.
- MALTZMAN, I., BROOKS, L. O., BOGARTZ, W., & SUMMERS, S. S. The facilitation of problem solving by prior exposure to uncommon responses. *Journal of Experimental Psychology*, 1958, 56, 399-406.
- MALTZMAN, I., SIMON, S., RASKIN, D., & LIGHT, L. Experimental studies in the training of originality. *Psychological Monographs*, 1960, 74(6, Whole No. 493).
- MEDNICK, M. T., MEDNICK, S. A., & MEDNICK, E. V. Incubation of creative performance and specific associative priming. *Journal of Abnormal and Social Psychology*, 1964, 69, 84-88.
- MEDNICK, S. A. The associative basis of the creative process. *Psychological Review*, 1962, 69, 220-232.
- MILTON, G. A. The effects of sex-role identification upon problem-solving skills. *Journal of Abnormal and Social Psychology*, 1957, 55, 208-212.
- POINCARÉ, H. Mathematical creation. In B. Ghiselin (Ed.), *The creative process*. New York: New American Library, 1952. Pp. 33-42.
- TYLER, L. E. *The psychology of individual differences*. (3rd ed.) New York: Appleton-Century-Crofts, 1965.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.
- WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.
- WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology* (Rev. ed.) New York: Holt, 1954.
- YOUTZ, R. P. Psychological foundations of applied imagination. In S. J. Parnes & H. F. Harding (Eds.), *A source book for creative thinking*. New York: Scribners, 1962. Pp. 193-215.

³ Maltzman et al. (1964) were subsequently unsuccessful in facilitating solutions to this problem.

SOME SOCIAL CLASS DIFFERENCES IN HELPING BEHAVIOR¹

LEONARD BERKOWITZ² AND PHILIP FRIEDMAN

University of Wisconsin

This experiment tested the hypothesis that members of the entrepreneurial middle class would be more likely than people from the bureaucratic middle class to act in accord with social exchange conceptions and display a reciprocity orientation toward giving and getting. 345 white adolescent boys were individually placed in an experimental situation in which they (a) first received high or low help when they were in need of assistance, and then (b) were given an opportunity to help someone, either the person who had worked for them or someone else. As expected, the later work in behalf of the other person was more greatly affected by the level of help received earlier in the entrepreneurial middle-class sample than in either the bureaucratic middle-class or working-class samples. The entrepreneurial boys were most apt to give only to the extent they had gotten. Effect of the other person's social class was also considered.

A variety of recent investigations have demonstrated that American college students often go out of their way to help other people who need their assistance even when they do not anticipate immediate return benefits. Such observations are at once consistent with and outside the realm of contemporary social psychological theorizing. Thus, in accord with these demonstrations of altruism, researchers have shown that much of human social behavior violates the simple incentive principles operative in animal behavior; human actions could be governed by considerations other than those for maximum self-gain. Where rats generally strive for maximum immediate rewards at minimum costs to themselves (Logan, 1960), human beings at times act in ways that, strictly speaking, are not altogether utilitarian. Homans (1961) has contended, for example, that many social relationships

are regulated by a desire for distributive justice. A person supposedly wants all of the people involved in any given social relationship to be equal in the ratio of their rewards to their investments. As a result, the individual theoretically wants, not as much as he can get from the relationship, but only as much as he believes he deserves. Similarly, Gouldner (1960) has argued that there is a universal, moral norm of reciprocity demanding, at the very least, that people should help and not injure those who have helped them. This normative expectation often inhibits exploitations, Gouldner points out, "safeguarding powerful people against the temptation of their own status [p. 174]." Instead of attempting to obtain the maximum possible rewards that his power can bring, the powerful individual who has adopted this norm supposedly feels obligated to reciprocate, to some extent, for the benefits he has received from those who are less powerful.

Within the limitations established by these normative expectations, however, social interaction is often regarded as being akin to a financial exchange. According to this reasoning, a person gives rewards to other people primarily in order to receive benefits from them in return, and typically acts as he does because he believes the action is to his benefit. He is supposedly guided largely by considerations of rewards and costs (cf. Thibaut & Kelley, 1959). From this point of view, giving help without at least anticipation of

¹The research reported in this paper was carried out under Grant GS-345 from the National Science Foundation to the senior author. Philip Friedman served as the experimenter under Leonard Berkowitz' supervision. We are grateful to Louise Daniels Epps for her assistance in the conduct of the experiment, to Philip Lambert of the University of Wisconsin, and to the principals of the Madison, Wisconsin, high schools for their kind cooperation in helping us recruit the subjects. Preliminary summaries of some of the findings were presented by Philip Friedman at the 1965 meetings of the Midwestern Psychological Association, and by Leonard Berkowitz in the *International Journal of Psychology*, edited in Paris by G. de Montmollin.

²Now at Cornell University.

return benefits from other persons would be somewhat surprising.

However, reciprocity and distributive justice expectations may not be the only rules controlling the degree to which social interactions are conducted as an economic exchange. The senior author and his co-workers (Berkowitz & Daniels, 1963, 1964; Daniels & Berkowitz, 1963; Goranson & Berkowitz, 1966) have suggested that middle-class American society (among other groups) possesses an ideal or norm prescribing that the individual should help other people who are dependent upon him and need his assistance. When someone in this society learns that a person needs his help, he presumably becomes aware of this "social responsibility norm," and feels obligated to aid this dependent person even when no direct return benefits are anticipated. In line with this suggestion, according to their research, middle-class American university students typically exert greater effort in behalf of a highly dependent peer than for a less dependent person even though (a) the dependent individual could gain a monetary prize and there was no prize for themselves, and (b) neither the dependent person nor the low status experimenter would supposedly find out about their work for at least a month. Moreover, when they were later asked why they had helped the person in need of their assistance, many of the subjects reported having felt obligated to give the help, and also said that other persons would have expected them to aid the individual who was dependent upon them. Outside the Wisconsin laboratory, Schopler has demonstrated that college students will give greater help to the person needing their aid the lower the cost of the assistance to themselves (Schopler & Bateson, 1965), and will give assistance when it is needed because of circumstances beyond the dependent individual's control (Schopler & Matthews, 1965).

But although these findings confirm the existence of the social responsibility norm in the American middle class, from which most of the university students had come, they also raise the question of the generality of this standard of conduct. Would working-class people also be willing to help someone

who was dependent upon them when there was no likelihood of return benefits? Social class differences probably exist in this mode of behavior. Simmel (1950) has suggested that economic exchange conceptions of social interaction are an outgrowth of experience with a money economy. Reasoning that members of the middle class would have a greater amount of this financial experience than would working-class people, Muir and Weinstein (1962) have reported that their middle-class female respondents were more inclined than the working-class women to deal with other people in exchange terms. Thus, a higher proportion of the middle-class women said they felt obligated to do favors "in return," while such received favors supposedly mattered less to the working-class women. Similarly, more of the high socioeconomic status women insisted they would stop doing favors for people who failed to repay their social obligations even though they were able to do so. The investigators concluded that the working-class sample was more likely to exhibit a "modified altruism, in that one gives when able . . .," while the middle-class women "consciously used an economic model, with business-like norms, in their sociable interaction [p. 538]." This conclusion suggests that the middle class is more strongly governed by *reciprocity* considerations involving ratios of giving and getting, and that the working class is more likely to be *responsibility oriented* so that less attention is given to the benefits received for an action.

Such a conclusion may be premature, however. The Muir and Weinstein findings may be limited to the particular procedure they followed. For one thing, attitude surveys of representative samples of adults in five nations found that admiration of generosity and considerateness as personal qualities generally increases with education (Almond & Verba, 1963, p. 265). Assuming that educational level is an approximate indicator of social class, this observation may mean that high socioeconomic status people typically are more likely than their lower socioeconomic status counterparts to hold the values conducive to social responsibility and altruism. Furthermore, other interview data collected in the

United States (Kohn, 1959) indicate that working-class adults are usually more concerned with the immediate consequences of their children's actions than are middle-class parents, while the latter place greater emphasis upon the inculcation of abstract moral principles. Social responsibility expectations are relatively abstract. Where a reciprocity orientation focuses upon the fairly immediate and concrete consequences of an action—"what's in it for me, and what will it cost me?"—responsibility ideals tell the individual he should conform to an abstract code of conduct, regardless of the immediate rewards and costs—"do what is right because it is right." To the extent that the society at large possesses these responsibility ideals, then, members of the middle class may be somewhat more likely to hold and adhere to these abstract standards of conduct than people from the working class.

But in addition, Simmel's (1950) analysis of social exchange as an outgrowth of monetary experiences clearly implies that the category, "middle class," can be too undifferentiated for many purposes. Many people in this group actually have relatively little experience with financial transactions aside from family and personal purchases. That segment of the middle class most actively involved in money matters and the exchange of goods and services should be most inclined to think of social interactions in economic terms. The distinction between "entrepreneurial" and "bureaucratic" middle class proposed by Miller and Swanson (1958) is relevant here. In comparison to the bureaucratic people, entrepreneurs are more concerned with the exchange of money and services, with giving and getting. The bureaucratic middle-class person, on the other hand, is theoretically more likely to behave in accord with the rules and regulations of his society. The impersonal organization within which he works customarily places great emphasis upon adherence to the standards of the organization, often regardless of personal considerations. The bureaucratic individual growing up in a society prescribing that people should help those who are dependent upon them should therefore be more willing than middle-class entrepreneurs to assist other persons in need

of aid regardless of the benefits he can derive or has obtained from the situation.

The dependent person's social class might also influence the magnitude of the help that is given him because of attitudes toward his status group. Daniels and Berkowitz (1963) showed that motivation to aid the dependent person is typically greater, the higher the liking for him. Two different trends must be considered in this regard. On one hand, the attractiveness of a person is often in direct relation to his social status (Grossman & Wrighter, 1948; Strodtbeck, James, & Hawkins, 1958). There may thus be a greater general willingness to help someone from the middle class than from the working class. However, to make a contrary prediction, social contacts and friendships are usually more frequent within a social class than across social class lines (Hollingshead, 1949). Many people may therefore be more inclined to aid another individual from the same social class as their own than someone from a different status level. Reluctance to assist someone from a different class background could well be indicative of feelings of estrangement from that class.

METHOD

Subjects

The subjects were 345 white students between 13 and 16 years of age from the three Madison, Wisconsin, high schools who had volunteered to serve in a psychology experiment for which they would be paid \$1.50.

Social class distinctions. The design called for a classification of the boys into three social class groups, based on their reports of their fathers' occupation and education: entrepreneurial middle class, bureaucratic middle class, and working class. The major distinction between the middle class and working class was in terms of "white-" versus "blue-collar" jobs, although in borderline cases attention was also given to the fathers' educational attainments. Entrepreneurs were defined as people whose income was determined to a considerable extent by their own independent activity, and included: (a) all owners of businesses, not including self-employed laborers, (b) all sales people, and (c) professionals working either for themselves or in partnership. Members of the bureaucratic middle class, on the other hand, derived their income mainly from working for someone else; the bureaucratic wage earner typically was employed in a multilevel organization. The occupations of the bureaucratic and entrepreneurial fathers of the subjects were rated employing the Hatt-North occupational prestige data (National

Opinion Research Center, 1953), and although the bureaucratic occupations tended to have somewhat higher prestige the difference was not significant ($F = 2.61$, $df = 1$, 185).

Procedure

The experiment was conducted in two phases. The experimenter first met groups of students, who had previously indicated their interest in participating in the experiment, in the students' school and administered several questionnaires to the boys. The items, dealing with family background and personal attitudes and interests, provided information permitting the social classification of the boys, and also supported one of the experimental manipulations. All of the boys were told the experiment sought to determine whether boys of their age could be effective supervisors of other youngsters' work. A simple and brief work situation supposedly would be established in which the boy would have to work for about an hour.

Some time later, a youngster was contacted and asked to come to the laboratory at a designated day and hour (after school). When he arrived, he was ushered into a separate room so that discussion with or even sight of the other subjects also scheduled at this time would be minimal. Each boy was reminded that the study had to do with the ability of adolescents to supervise other people and was informed that the study would be carried out in two unconnected parts. In the first session he would be the supervisor who had to guide the work of another student in a nearby room. He would be shown the task that the other boy, the worker (who actually was nonexistent), would have to perform, and would have to write instructions to the worker telling him how to carry out his job. Written notes were to be used, the experimenter said, because that was how many supervisors worked in industry today. The supervisor's task was to explain to the worker how to draw a given geometric design accurately and rapidly. The experimenter went on to say the boy would have 10 minutes to write his instructions, which the experimenter would then bring to the worker, and the worker would have 10 minutes to make as many of the designs as possible.

The boy was also informed that the supervisor earning the highest grade in this first part would receive a \$10 gift certificate, and that this grade would be based on two criteria: 80% of the grade would be determined by the number of designs drawn by the worker (a good supervisor has to be able to get his men to work hard, it was explained), and the rest would be based on how well he wrote the instructions. This information led the subject to believe that his likelihood of gaining the money prize was very greatly dependent upon his partner's work in his behalf. The money prize was then shown to the subject along with an example of the design to be drawn, and the subject was then left alone to write the instructions for his worker as well as a message urging the worker to work hard. Ten minutes later the note was picked up, suppos-

edly to be delivered to the worker. (In order to make sure that the subject would not blame himself for any low production by his partner, the experimenter looked at the subject's note and said the instructions were good.) At the end of another 10-minute interval, during which the subject was given an irrelevant task so as to keep him busy, the experimenter returned with the worker's productivity.

This new information constituted the first experimental manipulation. In one condition within each of the social class groupings, the experimenter showed the subject that his partner had made a good many of the required designs. He pointed out to the boy that his partner had helped him a great deal by working hard in his behalf (*high help received* condition). The remaining subjects within the social class groupings were shown that the partner had made only two designs, and the experimenter commented that the worker had given the boy very little help (*low help received* condition). With this information, the subject was told the first part of the experiment was over, and that there would now be an entirely new part.

In this second phase there would be an entirely different contest, the experimenter said. The subject now was to be the worker and someone else would be the supervisor who was eligible to win a different cash prize. Within each of the groups established up to this point, half of the subjects were led to believe they would be working for the person who had been their worker earlier (*same person* condition), while the other subjects were informed their supervisor would be a boy they had not encountered before (*different person* condition). Since this was a new contest, the experimenter continued, there would be somewhat different rules. The best supervisor would win another \$10 prize, but this time 70% of the supervisor's grade would be based on the subject's productivity.³ Thus, all subjects were made to believe their supervisors were very dependent upon them.

The final manipulation was then established. Telling the subject that he had already shown his supervisor the new job and that the supervisor was now writing his instructions, the experimenter asked the subject to indicate his first impressions of his supervisor's personality based on biographical information the experimenter would provide. The experimenter then gave the boy a biographical data sheet containing fictitious information about his supervisor and also a "personality sketch" questionnaire on which the subject was to indicate his impressions of this person. For half of the cases, the biographical information portrayed the supervisor as coming from a clearly *middle-class* family (thus, the boy's father supposedly was the manager of a small department store), while in the remaining cases the

³ The degree of dependency was changed from 80% to 70% in order to emphasize the fact that this was a new "contest," and that the supervisor's likelihood of gaining a prize in this second contest would not affect the subject's chances of winning money in the first contest.

supervisor was shown to be from the *working class* (among other things, his father was said to be an automobile repairman). In the *different person* condition, in addition to giving the subject this information about his second-phase supervisor, he was led to believe his first-phase worker had the same class background as his supervisor: the worker's father was therefore said to be either an office manager or an industrial worker.

After the information about the supervisor was provided, the experimenter returned with the supervisor's supposed instructions and the materials needed for the job. Following his instructions, the boy would have to cut out and paste together very simple specified designs from the paper supplied to him, making "paper moons on a paper sky." The subject was told not to mass produce but to work on only one unit at a time. Each unit, furthermore, was to be placed inside an envelope, and the envelope was to be sealed at the end of the work period. The experimenter explained that the envelope would not be opened for several months so that he would not know how hard the subject had worked and would make his ratings of the supervisor's instructions without being influenced by knowledge of the boy's performance. (This information was given to the subject in order to minimize his motivation to work hard in order to gain the experimenter's approval.) There would be a brief practice period, the experimenter continued, and then a 20-minute work period. Only the work done during this latter period would count. Approximately 8 minutes later the experimenter returned, told the boy the practice period was over and that the main work period was about to begin, and reminded him to place his products within the envelope and seal it. The number of units made by the subject, supposedly in his partner's behalf, constituted the measure of the subject's motivation to help the other boy in the absence of direct return benefits.

A final questionnaire was administered at the end of the 20-minute work period. When this was completed the boy was paid and was asked not to talk about the experiment.

RESULTS

Effectiveness of Experimental Manipulations

Almost every questionnaire item attests to the success of the *help received* manipulation. Analyses of variance of the scores on these postexperiment measures demonstrated that the subjects generally had a much more favorable attitude after learning they had received considerable help from their peer than after being told their first partner had given them little help. More important than this, however, are the results showing that the subjects had at least listened to the experimenter's statement as to who was their supervisor in the second phase. There were statistically significant interactions between the *help received* and *same person versus different person* variables on five of the postexperiment questionnaire and three of the "personality sketch" items, as indicated in Table 1. Thus, when the subjects had been informed that their supervisor was *different* from the person who had worked for them earlier, the level of help they had received in the first phase from the other person did not markedly influence their subsequent ratings of their supervisor on: (Item 1) the quality of their instruction, (Item 2) how much they wanted to get to know the supervisor better, (Item 5) whether or not they would vote for the supervisor if he were a candidate for a school office, (Item 8) the degree to which they thought they would have worked harder if their supervisor had been someone else, and (Item 9) their belief that their supervisor and they had "the

TABLE 1
ATTITUDES TOWARD SECOND-PHASE SUPERVISOR

Item	High help received		Low help received		Difference between help levels	
	Same person (N=92)	Different person (N=84)	Same person (N=84)	Different person (N=85)	Same person	Different person
1. Grade for supervisor's instructions	1.78	2.15	2.85	2.00	+1.07	-.15
2. Want to know supervisor better	2.35	2.80	4.15	2.86	+1.80	+.06
5. Vote for him for school office	3.35	3.90	5.68	4.01	+2.33	+.11
8. Would have worked harder for another supervisor	9.85	8.98	7.86	8.93	-1.99	-.05
9. He and I have same feelings	5.21	5.11	7.09	5.76	+1.88	+.65
PS-1. Is neat individual	3.21	3.93	4.63	4.30	+1.42	+.37
PS-3. More likable than average	4.18	4.84	5.84	5.18	+1.66	+.34
PS-10. Should hold office	3.48	4.56	5.73	5.24	+2.25	+.68

* Note.—For all of these items there was a statistically significant interaction (at .05 level or less) between the *help received* and *same-different person* variables. For each item, except 8, the higher the score the more unfavorable the attitude.

TABLE 2

MEAN SCORE ON QUESTION, "HOW MUCH WOULD OTHER PEOPLE HAVE EXPECTED YOU TO WORK HARD FOR YOUR SUPERVISOR?"

Help received	Social class		
	Bureaucratic	Entrepreneurial middle	Working class
High	3.03 _{ab} (63)	2.40 _a (52)	3.48 _{ab} (61)
Low	3.68 _{ab} (60)	4.39 _a (49)	3.61 _{ab} (60)

Note.—The scale runs from 1 (a great deal) to 11 (very little). Cells having a letter in common are not significantly different from each other at the .05 level of confidence by Duncan multiple-range test corrected for unequal cell frequencies. The number in parentheses is the number of cases in the given condition.

same feelings about things in life." All in all, as can be seen in the right-hand column of Table 1, their attitudes toward the *different person* supervisor were not greatly affected by the help the other person had given them. When the second-phase supervisor was the same person who had worked for them earlier, on the other hand, the level of help the subjects had received from this person influenced their subsequent ratings of him on these questions to a much greater extent. Generally speaking, the subjects were more favorably disposed toward the boy who had given them a great deal of help rather than little help. The subjects' social class membership did not significantly alter these findings.

Much the same significant interaction between the *help received* and *same-different person* variables was obtained with three of the personality-sketch items. This instrument was completed, it will be recalled, after the experimental variations were established, but before the subjects had to work for their supervisor. Relative to the fairly similar ratings given to the *different person* supervisor in the two *help* conditions, the youngsters regarded the *same* boy giving high prior help as being: (PS-1) personally neater, (PS-3) more likable, and (PS-10) more definitely the sort of person who should hold a school office, than the *same* youngster who had given them little help earlier.

Normative Expectations Regarding Help Giving

Although the subjects' social class was generally not an important determinant of the

above-mentioned attitudes toward the boy needing their assistance (with the exception of PS-3, likability), their class background did influence their perception of the social propriety of helping him. One of the items on the final questionnaire asked, "How much would other people have expected you to work hard for your supervisor?" This item presumably reflects the extent to which the subjects believed working hard in behalf of the supervisor was required by the normative expectations of most people in their environment. There were only two significant terms in the analysis of variance: the main effect for *help received*, and the interaction of *help received* and *class of subject*. The means involved in this latter significant interaction are given in Table 2. The *help received* effect clearly indicates that all of the boys recognized obligations to reciprocate; if they had gotten high help, they were expected to give high help in return. The strength of this expectation varied somewhat, however, among the social class groups. As can be seen in the table, the entrepreneurial middle-class boys were the ones who were most affected by the help they had gotten earlier. Thus, in keeping with the authors' theoretical expectations, the youngsters from entrepreneurial middle-class families were somewhat more prone to display a reciprocity orientation toward getting and giving help than were the bureaucratic middle-class adolescents or the working-class boys in the sample.

Productivity

The analysis of variance of the productivity scores yielded several significant terms for *help received*, *subject's class*, the interaction of *supervisor's class* and *subject's class*,⁴ and the interaction of *help received* and *subject's class*. This latter interaction demonstrates that the actual work carried out in behalf of the dependent peer parallels the

⁴ This significant interaction is of some interest in that it indicates the boys generally worked hardest for a supervisor from their own social class background. The bureaucratic middle-class youngsters had a mean productivity of 17.7 for a middle-class supervisor and only 16.4 for a working-class supervisor, while the working-class subjects had a mean productivity of 14.6 for a middle-class boy and 16.2 for a fellow working-class boy.

above mentioned normative expectations to some extent. Turning to the condition means shown in Table 3, it is apparent that the bureaucratic middle-class subjects worked significantly harder over all experimental conditions combined than the working-class youngsters. The bureaucratic middle-class and working-class boys, moreover, were not reliably affected by the level of help they had received earlier. By contrast, the entrepreneurial subjects gave their supervisor significantly greater help after receiving high help than after getting little help.

Effects of Other's Social Class Membership

There are a good many indications in the present study that the youngsters' attitudes toward their second-phase supervisor were influenced by knowledge of his social class. It should be kept in mind that the subjects in the *different person* condition had encountered two boys from the same social status level, and, consequently, their feelings toward the first-phase worker might have generalized to some degree to the second-phase supervisor because of this similarity in background.) Significant main effects for *other's social class* on a number of the questionnaire measures demonstrated that the middle-class boys typically were more attractive to both their lower- and middle-class peers than were youngsters from the working class. Thus, in comparison with the working-class boy, and over all subjects' classes, the middle-class teenager was significantly more definitely:

TABLE 3
PRODUCTIVITY MEANS IN CONDITIONS INVOLVED IN SIGNIFICANT "HELP RECEIVED BY SUBJECT'S CLASS" INTERACTION

Help received	Subjects			
	Bureaucratic middle class		Working class	
	High help	Low help	High help	Low help
High	17.37 _{ab}	16.75 _{ab}	15.51 _{abd} (61)	15.1
Low	16.75 _{ab}	14.74 _{ab}	15.51 _{abd} (60)	15.1
Total	17.08 _a	16.76 _a	15.44 _a	

Note.—The numbers in parentheses are the number of cases in the 2nd condition. Means having a subscript in common and significant difference at the .05 level by a two-tailed minimum range test corrected for unequal cell frequencies.

(PS-2) wanted in the subject's circle of close friends, (PS-3) regarded as being more likable than the average person, (PS-5) judged as being sophisticated, (PS-6) wanted in the social organization of greatest importance to the subject, (PS-7) judged as mixing easily with other people, (PS-8) predicted to be successful in school, (PS-9) judged as being less insecure, and (PS-10) wanted for a school office. Furthermore, on the postexperiment questionnaire administered at the end of the session, the subjects were also reliably more definite in saying their parents would like to meet the parents of the middle-class youngster.

In addition to this general trend in favor of the middle-class supervisor, there were three measures on which there was an interaction between the subject's own social class and the other youngster's class background.

TABLE 4

MEASURES YIELDING A SIGNIFICANT INTERACTION BETWEEN SUBJECT'S AND OTHER'S SOCIAL CLASS

Measure	S's class											
	Bureaucratic middle class				Entrepreneurial middle class				Working class			
	High help		Low help		High help		Low help		High help		Low help	
	MC other	WC other	MC other	WC other	MC other	WC other	MC other	WC other	MC other	WC other	MC other	WC other
Productivity	(30) 17.37 _{ab}	(33) 17.03 _{ab}	(30) 18.03 _a	(30) 15.87 _{abed}	(26) 17.42 _{ab}	(26) 18.46 _a	(24) 16.04 _{abed}	(25) 13.54 _d	(27) 15.01 _{abd}	(34) 16.00 _{abd}	(30) 14.27 _{cd}	(30) 16.50 _{ab}
Job boring	7.30 _{ab}	7.46 _{ab}	7.17 _{ab}	7.37 _{ab}	7.12 _{ab}	8.50 _a	7.75 _{ab}	5.84 _b	8.34 _a	7.70 _{ab}	7.33 _{ab}	8.67 _a
Supervisor likable	4.17 _b	4.89 _d	4.40 _{bc}	6.53 _g	4.19 _b	4.93 _d	4.79 _{cd}	6.17 _g	3.68 _a	5.21 _{de}	5.67 _f	5.50 _{ef}

Note.—The numbers in parentheses are the number of cases in the given condition. For Item 4, a high score indicates a favorable attitude toward the second-phase job, while for PS-3 a high score indicates that the second-phase supervisor was rated as definitely not "more likable than the average person." Both scales have a possible range from 1 to 11. For each measure considered separately cells having a subscript in common are not significantly different at the .05 level.

These scores were (a) the subject's productivity in behalf of the supervisor, (b) his rating of how boring the work had been, and (c) his rating of the supervisor's likability. Since the latter two ratings yielded significant interactions of *Other's Social Class* \times *Subject's Class* \times *Help Received*, tests were made of the significance of the differences among all of the means involved in this second-order interaction for each of the three measures. (For the productivity data, however, only the interaction of *other's social class* and *subject's social class* attained significance.) The results are given in Table 4.

Looking first at the work index, it can be seen that the only reliable differences within any of the social class groupings arose in the entrepreneurial middle-class sample. Those entrepreneurial boys who had received little help in the first phase worked significantly less hard for their supervisor than did their entrepreneurial counterparts who had gotten a good deal of help earlier—but only when the less-helped subjects were now dealing with a working-class supervisor.⁵ The entrepreneurial boys maintained a fairly high level of productivity after receiving little help when their present effort was in behalf of another middle-class youngster. The reciprocity displayed by the entrepreneurial subjects apparently was somewhat selective.

The next item, the subject's rating of how boring his job of worker had been, can perhaps be taken as another indication of task motivation. Making this assumption, we again see that the most task resistance was expressed by the entrepreneurial subjects who had received little help themselves earlier and were now working for a working-class supervisor.

These last-mentioned findings do not mean that the entrepreneurial middle-class subjects were the only ones who were attentive to the class background of the other boys they had encountered in the experiment. The last line on Table 4, reporting the mean likability ratings given the second-phase supervisor, shows that both the bureaucratic and en-

trepreneurial middle-class youngsters had something of a negative attitude toward boys from the lower status level. Within each of the *help received* conditions, and in both middle-class groups, the middle-class supervisor was regarded as definitely more likable than the working-class supervisor. (The working-class subjects who had received high help earlier were also reliably more favorably disposed toward the middle-class than the working-class supervisor.) Equally important, the bureaucratic middle-class boys had just as unfavorable an attitude toward the working-class supervisor after getting low prior help as did the entrepreneurial middle-class subjects. The latter, entrepreneurial boys evidently assisted their supervisor in accord with their attitude toward him, while the bureaucratic boys seemed to be somewhat more responsive to the social responsibility ideal than to their attitude.

DISCUSSION

The present data are generally quite consistent with the authors' theoretical expectations. More than either the bureaucratic or working-class boys, the youngsters from entrepreneurial middle-class families followed a social exchange conception of interpersonal relationships and gave help only to the extent that they themselves had received help earlier. This was particularly true, we have seen, when the person they could assist was from a lower status level.

This last-mentioned effect of the help recipient's social background is important, and can help one to understand the apparent reciprocity orientation displayed by the entrepreneurial subjects. The authors suggest that the bureaucratic and entrepreneurial boys entered the experimental situation with a relatively strong bias against working-class people. Subsequent reactions to events were then affected by this initial attitude, with the entrepreneurial subjects being more likely to act in accord with their feelings. If the entrepreneurial subjects had received a good deal of help from a working-class boy, they were pleasantly surprised and felt they should reciprocate high help for high help. Getting little assistance when they were in need, on the other hand, was annoying if not anger-

⁵ Their work for the lower-class boy was lowest when he was the same person who had given them little help, but was also very low when the supervisor was a different person.

TABLE 5

MEAN FRIENDLINESS TOWARD SECOND-PHASE SUPERVISOR JUST BEFORE WORK FOR HIM

Measure	Bureaucratic middle class				Entrepreneurial middle class				Working class			
	High help		Low help		High help		Low help		High help		Low help	
	MC other	WC other	MC other	WC other	MC other	WC other	MC other	WC other	MC other	WC other	MC other	WC other
S-2. Admit to . . . close friends	(30)	(33)	(30)	(30)	(26)	(26)	(24)	(25)	(27)	(34)	(30)	(30)
	3.73 _e	3.84 _e	3.87 _e	5.53 _b	2.86 _{ab}	3.92 _c	4.58 _d	5.57 _e	2.66 _a	3.16 _b	4.40 _d	4.37 _d
PS-6. Admit into . . . social organization	2.97 _{ab}	4.10 _{de}	4.30 _{ef}	5.27 _b	2.98 _{ab}	3.16 _{bc}	3.83 _d	4.45 _f	2.80 _a	3.36 _c	4.93 _g	4.43 _f

Note.—For both measures the higher the score the lower the expressed friendliness. The possible range is from 1 to 11. Cells having a subscript in common are not significantly different at the .05 level by Duncan multiple-range test.

producing. The presence of a person against whom a bias existed, such as the working-class supervisor, elicited strong hostility from these provoked entrepreneurial boys and resulted in (a) unfavorable ratings of this lower status person, (b) a lowered motivation to help him, and (c) a derogation of the task on which they were supposed to work. They were particularly inclined to reciprocate low help for low help, then, when the person requiring their assistance was not highly regarded. The bureaucratic middle-class subjects, by contrast, were evidently less prone to act toward the working-class peer in keeping with their negative attitude toward him. These boys were significantly more unfriendly toward the working-class supervisor after getting little help than after receiving high help, but their effort in behalf of the working-class supervisor was not significantly affected by the level of help they had received.

REFERENCES

- ALMOND, G. A., & VERBA, S. *The civic culture*. Princeton: Princeton University Press, 1963.
- BERKOWITZ, L., & DANIELS, L. R. Responsibility and dependency. *Journal of Abnormal and Social Psychology*, 1963, **66**, 427-436.
- BERKOWITZ, L., & DANIELS, L. R. Affecting the salience of the social responsibility norm. *Journal of Abnormal and Social Psychology*, 1964, **68**, 302-306.
- DANIELS, L. R., & BERKOWITZ, L. Liking and response to dependency relations. *Human Relations*, 1963, **16**, 141-148.
- GORANSON, R. E., & BERKOWITZ, L. Reciprocity and responsibility reactions to prior help. *Journal of Personality and Social Psychology*, 1966, **3**, 227-232.
- GOULDNER, A. W. The norm of reciprocity: A preliminary statement. *American Sociological Review*, 1960, **25**, 161-178.
- GROSSMAN, B., & WRIGHT, J. The relationship between selection, rejection, and intelligence, social status and personality amongst sixth grade children. *Sociometry*, 1948, **11**, 346-355.
- HOLLINGSHEAD, A. B. *Elmtown's youth*. New York: Wiley, 1949.
- HOMANS, G. C. *Social behavior: Its elementary forms*. New York: Harcourt, Brace, 1961.
- KOHN, M. L. Social class and parental values. *American Journal of Sociology*, 1959, **64**, 337-351.
- LOGAN, F. A. *Incentive: How the conditions of reinforcement affect the performance of rats*. New Haven: Yale University Press, 1960.
- MILLER, D. R., & SWANSON, G. E. *The changing American parent*. New York: Wiley, 1958.
- MUIR, D. E., & WEINSTEIN, E. A. The social debt: An investigation of lower-class and middle-class norms of social obligation. *American Sociological Review*, 1962, **27**, 532-539.
- NATIONAL OPINION RESEARCH CENTER. Jobs and occupations: A popular evaluation. In R. Bendix & S. M. Lipset (Eds.), *Class, status and power*. Glencoe, Ill.: Free Press, 1953. Pp. 411-426.
- SCHOPLER, J., & BATESON, N. The power of dependence. *Journal of Personality and Social Psychology*, 1965, **2**, 247-254.
- SCHOPLER, J., & MATTHEWS, M. W. The influence of the perceived causal locus of partner's dependence on the use of interpersonal power. *Journal of Personality and Social Psychology*, 1965, **2**, 609-612.
- SEMMEL, G. *The sociology of Georg Simmel*. Glencoe, Ill.: Free Press, 1950.
- STRODTBECK, F., JAMES, R., & HAWKINS, C. Social status in jury deliberations. In E. Maccoby, E. Hartley, & T. Newcomb (Eds.), *Readings in social psychology*. New York: Holt, Rinehart & Winston, 1958. Pp. 379-388.
- THIBAUT, J. W., & KELLEY, H. H. *The social psychology of groups*. New York: Wiley, 1959.

(Received May 18, 1966)

BRIEF ARTICLES

EFFECTS OF AMOUNT OF PRIOR SUCCESS AND FAILURE ON EXPECTATIONS OF SUCCESS AND SUBSEQUENT TASK PERFORMANCE¹

N. T. FEATHER AND MARGARET R. SAVILLE

University of New England, Australia

96 Ss worked at a 13-item anagrams test. Success and failure were manipulated in the first 3 items of the test in the 8 patterns possible: $\frac{1}{2}$ the Ss were told that the anagrams were easier than most (high expectation) and $\frac{1}{2}$ that they were more difficult than most (low expectation). All Ss rated their chances of success before attempting each anagram. The last 10 anagrams were of approximately 50% difficulty. Measures of *n* Achievement and test anxiety were available prior to task performance. Ratings of anxiety and disappointment were obtained from a postperformance questionnaire. Results showed that: (a) Probability estimates were modified by task performance, shifting more after predominant failure than after predominant success; (b) the small amount of prior experience had localized effects on subsequent performance; and (c) ratings of anxiety and disappointment were negatively related to task performance.

In a recent investigation (Feather, 1966) it was found that prior success and failure at a task influenced subsequent performance at the same task. Subjects who, by an experimental artifice, were induced to fail at the first 5 anagrams of a test consisting of 15 anagrams obtained lower performance scores on the final 10 anagrams of the test than did subjects who, by an experimental artifice, were induced to succeed at the first 5 anagrams. Effects of success and failure on subjects' estimates of probability of success were also investigated. As predicted, reported probabilities reflected the pattern of success and failure, increasing following success and decreasing following failure. These estimated probabilities were also found to shift more following the initial 5 failures than following the initial 5 successes.

The present investigation is both a replication and an extension of the preceding study. It seemed plausible that the effects noted above might be a function of the amount of prior success and failure. To explore this possibility prior experience was experimentally manipulated for the first 3 anagrams of a test consisting of 13 anagrams to obtain eight possible sequences ranging from uniform success on the first 3 anagrams, through a mixture of success and failure, to uniform failure on the first 3 anagrams. Hence, in contrast to the preceding study, amount of

prior success or failure was less (a maximum of three successes or three failures). Again, however, we were interested in the effects of this prior experience on subjects' reported probabilities of success and on their performance on the final 10 anagrams of the test.

METHOD

Subjects

Subjects were 134 undergraduate students (52 males, 82 females) from an introductory course in psychology at the University of New England in 1964. In the first of two testing sessions these subjects wrote stories to four pictures under neutral conditions using the standard procedure described by McClelland, Atkinson, Clark, and Lowell (1953) for obtaining stories.² The following four pictures were presented: 2, 48, 1, and 7 (using numbers assigned by Atkinson, 1958). In the same test session subjects completed the Test Anxiety Questionnaire (Mandler & Sarason, 1952). The projective protocols were scored for *n* Achievement by an experienced scorer (MRS) whose reliability had been previously established at .86 on the training materials (Smith & Feld, 1958). The Test Anxiety Questionnaire was scored as in previous investigations by assigning scores from 1 to 5 for each item in the direction of increasing anxiety and summing over items to obtain a total score. The mean *n* Achievement and test-anxiety scores for the 134 subjects were as follows: mean *n* Achievement = 4.38, *SD* = 4.97; mean test anxiety = 97.02, *SD* = 24.29. Both *n* Achievement and test-anxiety scores were trans-

¹ This study was conducted by the junior author as part of her honors program under the direction of the senior author.

² The first testing session was conducted by NTF.

TABLE 1
ANAGRAMS USED IN THE EXPERIMENT

Anagram used to induce		Final 10 anagrams (50% difficulty)	
Success	Failure		
RYAHD M-DEI VEREIA	AL-LEI OPUSON EMAGLE	WOELLY SPRUUE ONEASE ONEREP ONEGAH	OOTRUC UTTRUC MNEGAA MEEAIF HUNMER

lated to standard scores (i.e., to z scores). The difference between the n Achievement z score and the test-anxiety z score was calculated for each subject, that is, $z(n \text{ Achievement}) - z(\text{test anxiety})$.

Anagrams Task

Prior to the second testing session a pilot study was conducted to obtain 10 six-letter anagrams of between 40% and 60% difficulty with an average difficulty of about 50%. The anagrams selected according to these criteria appear in Table 1. Included in Table 1, also, are the anagrams used to induce the various patterns of prior success and failure. These anagrams were taken from those used in the original study (Feather, 1966) and were presented in sets of three in such a way that each subject experienced one of the eight possible combinations of success and/or failure on the first three trials of the test. For example, uniform failure was determined by presenting the three insoluble anagrams, uniform success by presenting the three very easy anagrams. The final 10 anagrams of approximately 50% difficulty were randomly ordered for each subject according to a table of random permutations (Cochran & Cox, 1957). The different sets of 13 anagrams were arranged in booklet form with one anagram per page.

The second testing session took place about 4 weeks after the initial experimental session.³ The test booklets were randomly distributed among subjects who were tested in small groups of 12-14 per group. All subjects received the following initial instructions:

The test that you are about to perform is a test of verbal intelligence. Please try to do your best as your scores will be taken as a fair and accurate indication of your intelligence level, particularly your ability to deal with verbal materials. The test consists of a set of disarranged words (anagrams). Your task is to rearrange each group of letters so that they make a meaningful English word. You will have 30 seconds to work at each anagram. Start when you are so instructed. Stop at the stop signal. Do not turn over a page until you are told to do so.

³ The second testing session was conducted by MRS.

Instructions for subjects in the high expectation condition then stated: "You should find these anagrams more difficult than most. About 70% of previous Psychology I students were able to solve them correctly in the time allowed." Subjects in the low expectation condition were told: "You should find these anagrams more difficult than most. About 30% of previous Psychology I students were able to solve them correctly in the time allowed."

Probability Estimates

Immediately before attempting each anagram and before seeing the anagram in sets were required to estimate their own chances of solving the anagram. This they did by using a 5-inch scale numbered from 0 to 100 in equal steps of 20, with the statement "No chance at all" at one extreme of the scale, the statement "An even chance" at the middle of the scale, and the statement "Completely certain" at the other extreme of the scale. A complete record of probability estimates throughout task performance was therefore available for each subject.

Postperformance Questionnaire

When they had attempted all 13 anagrams, subjects completed a postperformance questionnaire which included a Likert-type question concerned with degree of worry or anxiety experienced during task performance (scored 1-5 in the direction of increasing anxiety), a scale numbered from 0 to 15 which subjects checked to show where they felt most worried or anxious during task performance (scored as the ordinal position of the anagram where most worry and anxiety was reported), and a 5-inch scale on which subjects checked how disappointed they felt about task performance (scored 0-100 in the direction of increasing disappointment).

Composition of Groups

The eight different sequences of success and/or failure on the first three trials in combination with the two levels of reported difficulty of the task (high expectation versus low expectation) determined 16 different experimental conditions. An attempt was made to equate the 16 groups with respect to sex distribution and motive distribution. Each group consisted of either four females and two males or three females and three males. Similarly the z -score differences (n Achievement - test anxiety) in each group were roughly equivalent in distribution. Minor variations between the 16 groups with respect to sex and motive distribution did occur, but in the overall analysis of results these were of little consequence since the important comparisons involved combined groups with fairly large N s. As a result of this attempt to equate groups and in order to achieve equal N s in each of the 16 groups, the records of 24 subjects were deleted. The records of 14 Asian students were also excluded because of possible language difficulty. There remained 96 subjects, 6 in each experimental group.

TABLE 2
MEAN ESTIMATES OF PROBABILITY OF SUCCESS FOR ANAGRAMS OVER 13 TRIALS

Condition	Initial experience	Trials													Row mean
		1	2	3	4	5	6	7	8	9	10	11	12	13	
A	+++	.57	.74	.82	.86	.70	.67	.68	.66	.63	.65	.55	.50	.56	.66
B	++-	.64	.72	.78	.60	.59	.52	.49	.48	.44	.47	.47	.45	.44	.55
C	+-+	.62	.74	.45	.62	.59	.43	.47	.48	.56	.48	.52	.54	.45	.51
D	---	.56	.45	.53	.54	.55	.53	.52	.50	.55	.54	.55	.54	.56	.53
E	++-	.57	.39	.26	.37	.36	.36	.33	.35	.28	.31	.35	.33	.35	.33
F	+-+	.66	.48	.63	.50	.54	.53	.54	.55	.58	.55	.57	.57	.55	.56
G	+++	.53	.62	.47	.38	.44	.38	.33	.33	.33	.36	.39	.32	.39	.41
H	---	.56	.42	.34	.27	.26	.24	.21	.20	.19	.19	.21	.29	.28	.27
Column means		.59	.57	.53	.52	.50	.46	.44	.44	.44	.44	.45	.44	.44	

Note.—A + denotes success; a - denotes failure. Initial experience refers to the pattern of success and/or failure on Trials 1, 2, and 3.

RESULTS

Analysis of Probability Estimates

Mean change over trials. Table 2 presents means of the probability estimates obtained from subjects prior to attempting each of the 13 anagrams. These means are tabulated in relation to initial experience and trials. To conserve space the initial expectation conditions (high versus low) are not included in the classification.

An analysis of variance (Winer, 1962) shows that the main effect of initial experience is highly significant ($F = 7.70$, $df = 7/80$, $p < .001$). The overall mean probability estimates differ significantly across conditions ranging from .66 in Condition A (+++) to .27 in Condition H (---). Apparently initial experience had a pronounced effect on probability estimates. The main effect of trials is also highly significant ($F = 14.15$, $df = 12/960$, $p < .001$). Probability estimates averaged over all conditions show a tendency to decrease initially from a mean value of .59 and then, after five trials, to level out at a value of about .44. In addition, the interactive effect of initial experience and trials is highly significant ($F = 3.28$, $df = 84/960$, $p < .001$). Subjects tended to raise probability estimates after success and to lower them after failure. This can be clearly seen by examining changes in probability estimates from Trial 1 to Trial 4 across the eight different conditions where initial success and/or failure was controlled. All of these results may be taken to indicate the dominant influence of success and failure in modifying expectations of success. They strongly replicate the results of the previous study (Feather, 1966).

Probability estimates prior to Trial 1. Prob-

ability estimates obtained prior to performance at the first anagram tend to be higher in the high expectation conditions ($M = .61$) than in the low expectation conditions ($M = .56$). However, the difference between these means is not statistically significant. This difference was barely significant in the original study ($p < .05$). In future studies reported norms more divergent than 70% and 30% may have to be used to determine unequivocal differences in initial probability estimates. As in the preceding study the mean for the low expectation conditions suggests that initial expectations of success may have been closer to a P_s of .50 in these conditions than very low. Also, there is again a tendency for subjects to state initial estimates higher than the low reported norm (30%) and lower than the high reported norm (70%).

Absolute changes in probability estimates. The data in Table 2 suggest a tendency for probability estimates to change more after failure than after success. An analysis of absolute differences between probability estimates for Trial 1 and Trial 4 shows that the mean change (14.31) for Conditions A, B, C, and D where success predominates on the first three trials is significantly less than the mean change (21.58) for Conditions E, F, G, and H where failure predominates on the first three trials ($F = 6.10$, $df = 1/94$, $p < .05$). These data therefore agree with those of the original study in suggesting that expectations of success changed more after failure than after success, although the differences are less clear-cut than in the original study.⁴

⁴When initial probability estimates tend to be above .50, as in the present and original studies, the maximum possible increase in estimates following

Analysis of Performance Data

Mean number of anagrams solved. Table 3 presents the mean number of anagrams correctly solved out of the final 10 anagrams by subjects in the 16 experimental conditions, together with an analysis of variance of the data.

Table 3 fails to reveal any significant effects. This is in marked contrast to the results of the earlier study which indicated that performance scores were lower following uniform failure on the first five trials of the test than following uniform success on the first five trials of the test. On the assumption that the smaller amount of prior success and/or failure involved in the present study may have had a more localized effect on subsequent performance, a further analysis was conducted. Performance scores on the first 10 anagrams were partitioned into scores for the first 5 anagrams (Block 1) and scores for the last 5 anagrams (Block 2). These scores were analyzed for those subjects who had experienced uniform success (Condition A) and uniform failure (Condition H) and in relation to the initial expectation conditions (high versus low) so as to permit a comparison with the results of the earlier study where uniform success and failure were also involved. The results of this analysis are presented in Table 4 which also contains an analysis of the corresponding data from the original experiment (Feather, 1966).

Table 4 shows that the interactive effect of initial experience and blocks is statistically significant in the present study. Subjects who have experienced uniform success on the first 3 anagrams and then move to the relatively harder set of 10 anagrams of approximately 50% difficulty tend to obtain higher performance scores in Block 1 than in Block 2. Subjects who have experienced uniform failure on the first 3 anagrams and then move to the relatively easier set of 10 anagrams of approximately 50% difficulty tend to obtain higher performance scores in Block 2 than in Block 1. Initial success apparently determines an initial enhancement in performance followed by a decline as subjects work at what are relatively harder anagrams to them. In contrast, initial failure determines an initial depression in performance followed by an improve-

ment as subjects work at what are relatively easier anagrams to them. These results therefore indicate that prior experience had a marked effect on subsequent performance in the present study. However, Table 4 also shows that no such effects appear when a similar analysis is performed on the corresponding data of the original study. Apparently both the depressing effect of initial failure and the enhancing effect of initial success on subsequent performance are more widespread when the amount of prior success and failure is greater, as it is in the original study.

Analysis of Postperformance Data

Ratings of disappointment and anxiety. An analysis of variance applied to the *disappointment ratings* shows that the main effect of initial experience is statistically significant ($F = 2.30$, $df = 7/80$, $p < .05$). Mean disappointment ratings tend to be higher when the prior experience involves predominant failure than when it involves predominant success. Mean disappointment ratings are also higher in the high expectation condition than in the low expectation condition although the main effect of initial expectation is not statistically significant.

An analysis of variance applied to the *anxiety ratings* shows that the main effect of initial experience is statistically significant ($F = 3.80$, df

TABLE 3
MEAN NUMBER OF ANAGRAMS CORRECTLY ANSWERED
OUT OF FINAL 10 ANAGRAMS

Condition	Initial experience	Initial expectation of success		
		High (70% norm)	Low (30% norm)	Row means
A	+++	4.50	4.33	4.42
B	++	3.83	4.83	4.33
C	+-	5.00	3.33	4.17
D	-++	5.17	4.83	5.00
E	---	2.67	5.50	4.09
F	+-	4.83	5.50	5.17
G	+-	4.33	3.33	3.83
H	---	3.00	3.50	3.25
Column means		4.17	4.39	

Analysis of variance

Source	df	MS	F
Initial expectation (N)	1	.126	<1
Initial experience (E)	7	.452	<1
N × E	7	.567	1.04
Error	80	.544	

success (i.e., to +1.00) is less than the maximum possible decrease following failure (i.e., to 0). This difference in range of possible change might underlie the observed tendency for changes in expectation to be greater following failure than following success. The results would be more convincing if all initial estimates were at .50, but this would be difficult to arrange experimentally.

TABLE 4

MEAN NUMBER OF ANAGRAMS CORRECTLY ANSWERED IN BLOCK 1 (TRIALS 4-8) AND BLOCK 2 (TRIALS 9-13) IN CONDITIONS A AND H OF THE PRESENT STUDY AND IN BLOCK 1 (TRIALS 6-10) AND BLOCK 2 (TRIALS 11-15) FOR THE INITIAL SUCCESS AND INITIAL FAILURE CONDITIONS OF THE ORIGINAL STUDY (FEATHER, 1966)

	Condition	Initial expectation of success				
		Initial experience	Block	High (70% norm)	Low (30% norm)	Row means
Present study	A	+++	1	2.67	2.50	2.59
	H	---	2	1.83	1.83	1.83
			1	1.00	1.33	1.17
			2	2.00	2.17	2.09
		Column means		1.88	1.96	
Original study	Initial success	++++	1	2.94	3.00	2.97
	Initial failure	-----	2	2.94	3.28	3.11
			1	2.61	2.39	2.50
			2	1.83	2.56	2.20
		Column means		2.58	2.81	

Analysis of variance

Source	Present study (N = 24)			Original study (N = 72)		
	df	MS	F	df	MS	F
Between Ss	23			71		
Initial expectation (N)	1	.084	<1	1	1.778	<1
Initial experience (E)	1	4.084	1.60	1	17.362	8.41**
N × E	1	.332	<1	1	.027	<1
Error (between)	20	2.558		68	2.065	
Within Ss	24			72		
Blocks (B)	1	.084	<1	1	.250	<1
N × B	1	.000	<1	1	3.361	3.06
E × B	1	8.332	6.54*	1	1.777	1.62
N × E × B	1	.085	<1	1	1.001	<1
Error (within)	20	1.275		68	1.097	

* $p < .025$.
 ** $p < .01$.

= 7/80, $p < .005$). Mean anxiety ratings tend to be greater when the prior experience involves predominant failure than when it involves predominant success. The main effect of initial expectation is also statistically significant ($F = 5.62$, $df = 1/80$, $p < .025$). Subjects in the high expectation condition tend to report higher anxiety about performance than do subjects in the low expectation condition. Finally, the interactive effect of initial experience and initial expectation is statistically significant ($F = 2.92$, $df = 7/80$, $p < .01$). Prior experience involving predominant failure is especially likely to determine high anxiety ratings when the task is presented as relatively easy (high expectation condition).

Position of maximum anxiety. Subjects whose prior experience involved three successes (Condition A) report feeling most worried and upset about their performance near the ninth anagram attempted ($M = 8.79$). Subjects whose prior experience involved three failures (Condition H) report feeling most worried and upset about their performance near the fifth anagram ($M = 5.04$). The difference in means is statistically significant ($F = 9.05$, $df = 1/20$, $p < .01$). This difference may be related to the data in Table 4. Initial success is associated with lower mean performance in Block 2 than in Block 1 of the anagrams, and the position of maximum anxiety occurs in Block 2. Initial failure is associated with lower mean performance in Block 1 than in Block 2, of

the anagrams, and the position of maximum anxiety occurs in Block 1. Apparently prior failure results in both higher anxiety about performance and a more rapid development of anxiety about performance than does prior success. These differences in anxiety ratings correspond to differences in subsequent performance, high anxiety ratings being associated with low performance.

Interrelationship of Variables

Table 5 presents the intercorrelations of the scores involved in the preceding analyses for the entire group of 96 subjects.

These results will not be considered in detail. They are consistent with relationships that have been presented and discussed previously (Feather, 1963a, 1963b, 1965a) and with the results of the analyses that have been reported above. By way of summary it should be noted that:

1. There is a high positive correlation between probability estimates for Trial 4 and amount of prior success on Trials 1-3. Probability estimates for Trial 13 are positively related to number of anagrams correctly solved out of the final 10 anagrams. Both of these results attest to the modifying effect of task performance on reported probabilities.

2. Disappointment ratings, anxiety ratings, and test-anxiety scores are all positively interrelated. These are all self-report measures of emotional states.

3. Both disappointment ratings and anxiety ratings are negatively related to amount of prior success, to probability estimates for Trials 4 and

13, and to performance on the last 10 anagrams. It appears that amount of failure is a potent influence upon all of these ratings. As amount of prior failure increases, probability estimates tend to decline, ratings of anxiety and disappointment tend to increase, and maximum anxiety is reported earlier.

4. Test-anxiety scores and *n* Achievement scores are independent as in previous studies (see Atkinson & Feather, 1966).⁵

Discussion

The results of the present study replicate previous findings concerning changes in probability estimates following success and failure, and factors influencing reported anxiety and reported disappointment (Feather, 1963a, 1963b, 1965a). They again show that probability estimates and, by implication, expectations of success tend to increase after success and to decrease after failure and that, under the controlled conditions of the present study, they tend to change more after predominant failure than after predominant success. There is again ample evidence that reported anxiety and reported disappointment about performance both tend to increase as the amount of failure increases, and that these rat-

⁵ In the high expectation condition *n* Achievement scores and performance on the last 10 anagrams are negatively related ($r = -.338$, $df = 46$, $p < .02$), but unrelated in the low expectation condition ($r = .036$, $df = 46$, ns). This negative relationship, though not fully understood, was implied in discussion of another set of data (Feather, 1965b, p. 124).

TABLE 5
INTERCORRELATIONS OF SCORES ($N = 96$)

Variable	1	2	3	4	5	6	7	8	9	10
1. Probability estimate (Trial 1)		.396	.231	-.000	.010	.023	-.096	.003	.156	.066
2. Probability estimate (Trial 4)			.488	.021	-.033	-.320	.138	-.246	.145	.720
3. Probability estimate (Trial 13)				-.019	-.000	-.503	-.181	-.535	.639	.318
4. <i>n</i> Achievement					.153	.222	-.009	.196	-.164	-.056
5. Test anxiety						.247	-.060	.219	.098	-.078
6. Anxiety rating							.142	.658	-.356	-.336
7. Position of maximum anxiety								.210	-.306	.271
8. Disappointment rating									-.544	-.259
9. No. of anagrams correct (out of last 10)										.123
10. No. of prior successes										
<i>M</i>	58.53	51.42	43.89	4.77	96.08	2.91	5.79	61.40	4.28	1.50
<i>SD</i>	15.30	23.11	26.67	5.19	24.16	1.16	3.42	29.03	2.30	0.87

Note.—*r*'s of .200 and above are significant at $p < .05$ (2-tailed); *r*'s of .260 and above are significant at $p < .01$ (2-tailed); *r*'s of .329 and above are significant at $p < .001$ (2-tailed).

ings tend to be higher in the high expectation condition than in the low expectation condition.

The interesting new finding that emerges from the present study is that the effects of prior success and/or failure on subsequent performance may be localized. This is quite clear in Table 4 when one compares the means of the present study with those of the original study. When prior experience involves either three successes or three failures, as in the present study, the effects on subsequent performance appear to be limited to either the first or the second block of five anagrams. When prior experience involves either five successes or five failures, as in the original study, the effects on subsequent performance are more widespread extending over both blocks of anagrams. The influence of prior success or failure on subsequent performance becomes more limited as the amount of prior success or failure decreases.

The depressing effect of prior failure on subsequent performance may be conceived within the context of a *variable-incentive* model—as in the theory of achievement motivation (see Feather, 1966)—or a *fixed-incentive* model—implying a reduction in the resultant tendency to perform the task due to a progressive increase in the inhibitory tendency (T_{-} ; see Feather, 1963a, 1965a). Depressed performance may also follow because prior failure does not provide appropriate conditions under which *habits* necessary for subsequent success may be acquired (cf. concept of *response availability*, Atkinson & Feather, 1966). The effects of failure may also depend upon a buildup in inertial tendencies (Atkinson & Cartwright, 1964; Atkinson & Feather, 1966) that augments the resultant tendency to perform the task to such an extent that performance declines in accordance with the Yerkes-Dodson Law (Atkinson & O'Connor, 1966). Future investigations should be designed to test these overlapping theoretical interpretations that have developed within an expectancy-value framework.

The results of both the present study and the original study have clear implications for examination strategies. In a test involving similar items of varying difficulty it would seem advantageous to attempt easy items first in order to guarantee initial success. An unfortunate choice

of difficult items early in the test may result in initial failure and a detrimental effect on subsequent performance.

REFERENCES

- ATKINSON, J. W. (Ed.) *Motives in fantasy, action, and society*. Princeton, N. J.: Van Nostrand, 1958.
- ATKINSON, J. W., & CARTWRIGHT, D. Some neglected variables in contemporary conceptions of decision and performance. *Psychological Reports*, 1964, 14 (2), 575-590.
- ATKINSON, J. W., & FEATHER, N. T. (Eds.) *A theory of achievement motivation*. New York: Wiley, 1966.
- ATKINSON, J. W., & O'CONNOR, P. Neglected factors in studies of achievement-oriented performance: Social approval as incentive and performance decrement. In J. W. Atkinson & N. T. Feather (Eds.), *A theory of achievement motivation*. New York: Wiley, 1966. Pp. 299-325.
- COCHRAN, W. G., & COX, G. *Experimental designs*. New York: Wiley, 1957.
- FEATHER, N. T. The effect of differential failure on expectation of success, reported anxiety, and response uncertainty. *Journal of Personality*, 1963, 31, 289-312. (a)
- FEATHER, N. T. Mowrer's revised two-factor theory and the motive-expectancy-value model. *Psychological Review*, 1963, 70, 500-515. (b)
- FEATHER, N. T. Performance at a difficult task in relation to initial expectation of success, test anxiety, and need achievement. *Journal of Personality*, 1965, 33, 200-217. (a)
- FEATHER, N. T. The relationship of expectation of success to need achievement and test anxiety. *Journal of Personality and Social Psychology*, 1965, 1, 118-126. (b)
- FEATHER, N. T. Effects of prior success and failure on expectations of success and subsequent performance. *Journal of Personality and Social Psychology*, 1966, 3, 287-298.
- MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *Journal of Abnormal and Social Psychology*, 1952, 47, 166-173.
- MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
- SMITH, C. P., & FELD, S. C. How to learn the method of content analysis for *n* Achievement, *n* Affiliation, and *n* Power. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton, N. J.: Van Nostrand, 1958. Pp. 685-818.

(Received March 8, 1966)

INCREASED SELF-ACCEPTANCE: A MEANS OF REDUCING PREJUDICE¹

IRWIN M. RUBIN²

Alfred P. Sloan School of Management, Massachusetts Institute of Technology

An experiment was conducted to investigate the relationship between changes in an individual's level of self-acceptance and his level of ethnic prejudice. The 50 participants of the Osgood Hill summer program in sensitivity training made up the experimental population. $\frac{1}{3}$ of this population, randomly selected, served as an own control group. On the basis of these before after data, it was shown that: (a) Significant increases in self-acceptance and decreases in prejudice result from sensitivity training; (b) a significant positive relationship exists between changes in self-acceptance and changes in prejudice.

¹ In the area of ethnic prejudice, no single piece of research has had a greater impact than *The Authoritarian Personality* (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950).³ Frenkel-Brunswik, in summarizing the major findings of that research, stated:

Regardless of whether the specific topic was that of ambivalence, or aggression, or passivity, or some other related feature of personality dynamics, the outstanding finding was that the extremely unprejudiced individual tends to manifest a greater readiness to become aware of unacceptable tendencies and impulses in himself. The prejudiced individual, on the other hand, is more apt not to face these tendencies openly and thus to fail in integrating them satisfactorily with the conscious image [phenomenal self] he has of himself [p. 474].

From this naturally follows the hypothesis that one way to change a person's ethnic attitudes is to somehow alter his attitudes toward himself.³ Successful alteration of one's self-concept, however, requires a particular kind of environment—one in which the individual feels psychologically safe.

Sensitivity or laboratory training is one example of a change strategy which provides such an island of psychological safety.⁴ In a

broad sense, it can be defined as

... an educational strategy which is based primarily on the experiences generated in various social encounters by the learners themselves, and which aims to influence attitudes and develop competences toward learning about human interactions [Schein & Bennis, 1965, p. 4].

The high level of mutual trust which develops during the training-group sessions enhances an individual's willingness to openly explore aspects of his self-concept that before were kept hidden. The individual becomes more accepting⁵ of these previously guarded aspects of his self not only because the training-group members (as well as the trainer) reward his openness by being supportive, but, in addition, each individual comes to realize that his feelings are not unique. As Weschler (1962, p. 42) states:

Increasing experience suggests a great universality with respect to many closely guarded and deeply experienced feelings. We therefore attempt, in training, to help people become aware of their major areas of pluralistic ignorance. Thus it becomes possible for them to realize that, in fact, they no longer need to hide nearly so much of themselves from others as they formerly did.

In light of this discussion, the following specific hypotheses will be tested:

1. The higher an individual's level of self-acceptance,⁶ the lower will be his level of prejudice.

2. As a result of participation in a sensitivity training laboratory, an individual's level of self-acceptance will increase and his level of prejudice will decrease.

3. Changes in self-acceptance will be associated with changes in prejudice.

⁵ See Gordon (1950) and Bunker (1965) for data supporting the proposition that training groups increase self-acceptance.

¹ The research reported in this paper represents a part of the author's doctoral dissertation (Rubin, 1967) which was supported by a Ford Foundation fellowship in business administration. The author is grateful to the members of his committee, Edgar Schein, Warren Bennis, and David Berlew, all of MIT, and to Robert Kahn for their assistance in the dissertation.

² Special thanks must go to the entire staff of the Department of Human Relations at Boston University for their cooperation in the successful completion of this study.

³ A more complete and detailed review of prior research which supports this hypothesis can be found in Rubin (1966, 1967), Rogers (1951), and Wylie (1961).

⁴ For a complete discussion of all that is involved in a sensitivity training program see Schein and Bennis (1965).

METHOD

Definition of Self-Acceptance

Self-acceptance, as it will be used in this paper, involves a willingness to confront ego-alien as well as ego-syntonic aspects of the self and to accept rather than deny their existence. The Dorris, Levinson, and Hanfmann (1954) Sentence Completion Test was used to measure the level of an individual's self-acceptance.

The Sentence Completion Test comprises 50 sentence stems. Half the stems use first-person pronouns and half a third-person pronoun or proper name. First- and third-person items are matched in content and are randomly distributed rather than appearing sequentially. The only published data concerning reliability come from a study by Jones (1960) who found no significant changes on the Sentence Completion Test among the members of his matched control group after a period of 9 months.

The measure of self-acceptance used in this study was derived in the following manner. Individual stem completions were coded⁶ for ego-threatening content.⁷ An ego-threatening completion was defined as: "Any item which states or strongly implies any attitude, feeling, or action, which if accepted by⁸ . . . as applying to oneself, would involve confronting at least a mild degree of psychological pain." For example, expression of fears, socially unacceptable responses, admission of inferiority or incompetence, extreme hostility or aggression, etc., were coded as threatening.

The assumption was then made that the more willing a person is to admit the personal relevance of ego-threatening material, the greater his level of self-acceptance. Therefore, the number of ego-threatening responses next to which the respondent placed a (+), signifying that the completed sentence had some personal relevance, divided by the total number of ego-threatening responses (#ET) yields the measure of self-acceptance, *ETA*, used in this study. Scores ranged from 0-100% with a low score indicating a low level of self-acceptance.

⁶ Each pair of items was copied on a separate piece of paper with the respondent's identification number placed on the reverse side. This procedure made it impossible for the scorers to know if the response was "pre" or "post" and eliminated any halo effect that might have been created by reading an individual's total record.

⁷ The correlation coefficient between two samples, independently coded, was .89 (see Johnson, 1949, p. 97, for the formula used to compute this coefficient). The author gratefully acknowledges the assistance provided by his colleague, Douglas T. Hall, in this phase of the study.

⁸ For the females, the phrase, "the majority of women associated with the nursing profession," was inserted because virtually all the females in the experimental population fell into that category. For the males, who were more heterogeneous, the phrase, "the average male in our culture," was inserted. Two forms, male and female, of the scale were used for this research.

It is necessary at this point to state one further hypothesis: the absolute number of ego-threatening statements will not change as a result of sensitivity training. This means that *ETA* will increase because the number of ego-threatening statements accepted will increase and not because the number of ego-threatening statements decreases. The rationale here is that sensitivity training will not rid a person of his basic conflicts and anxieties nor does it attempt to help him make light of his times of crises. Instead, in some ideal sense it may help a person find in himself the natural tools which will enable him to effectively cope with these conflicts.

Definition of Prejudice

Prejudice, as it will be used in this paper, refers to the extent of an individual's willingness to accept others in terms of their common humanity, no matter how different they may seem from himself (Harding & Schuman, 1961). It can be characterized by such emotional manifestations as love, brotherhood, and sympathy.

The scale is made up of 15 items⁹ of the following type:

The white school board in a community builds two new schools and fixes the school lines so that almost all the colored children go to one new school and all the white children to the other new school. How do you suppose most of the Negroes in the community would react to this?

—(a) While there are some exceptions, many Negroes are mainly concerned with getting money for food, rent, and other things, and do not have too much interest in the matter of schools one way or the other.

—(b) Every community is different, and it is almost impossible for someone not living there to know enough about the situation to judge.

—(c) The average Negro mother or father would not like what the school board has done about school lines.

—(d) The average Negro parent would simply be pleased to have a new school for their children, especially if it was equal to the white school in every way.

The measure of prejudice used in this study was derived in the following manner. The respondent was asked to rank each of the four choices following an item from 1 ("most likely reaction") to 4 ("least likely reaction"). Each respondent's series of ranks was then compared to a theoretically ideal set of ranks,¹⁰ and the

⁹ Schuman and Harding (1963) report that the split-half reliability on an 11-item version is .76, while the retest reliability over a 1-month interval is .80. In addition, four control items are included to check on the extent to which response set is operating.

¹⁰ Howard Schuman and the writer independently ranked all items as to how the "least prejudiced person" would assign his ranks. We agreed on 100% of the first and second ranks and on 88% of the third and fourth ranks, yielding an overall percentage of agreement of 94%.

absolute difference between ranks was computed. The sum of these differences across the 15 experimental items yielded the respondent's prejudice score (*PR*). This score could range from 0-120 (i.e., 15 items \times a maximum difference of 8 points for any item) with a high score indicating a high level of prejudice.

Sample Characteristics

The delegates to the Osgood Hill summer program in sensitivity training in Andover, Massachusetts, constituted the laboratory population studied in this research. The program was 2 weeks in length (June 25-July 7), and the participants "lived in" in the sense that they slept on the premises and ate virtually all their meals together.

There were 50 delegates to the laboratory, 30 females and 20 males. They ranged in age from 23 to 59 with a mean age of 33 years. The majority had at least a BS degree and a few had advanced degrees. The majority came from the New England area, but several came from Miami, Cleveland, and Chicago. There were 8 Negroes in the population, and the trainers made certain that each of the five training groups which were formed had at least one Negro and an even proportion of males and females. Occupational groups represented are shown in Table 1.

Experimental Design and Procedure

The total available experimental group ($N = 50$) was randomly split into two groups of unequal size. The smaller group ($N = 14$) was tested by means of mail questionnaires 2 weeks prior to their arrival at Osgood Hill. The entire group was then tested upon their arrival, but before the first training-group session, and again the morning of the next to the last day of the laboratory. In other words, those people we have called controls also participated in the laboratory, but had been tested twice prior to the first training-group session. Though these controls participated in the training-group session, their postexperimental scores were not used in the analysis of results. This design can be depicted in the following manner:

June 11	June 25	July 5
Before controls	After controls	
	Training groups	
	Before	After
	experimentals	experimentals

Of the available control group of 14, 2 people never arrived and 1 returned an unusable questionnaire, leaving a final control group of 11. Of the available experimental group of 36, 1 missed the pretest and 5 returned unusable questionnaires, leaving 30 for the final experimental group.

RESULTS

Initial Correlation between ETA and PR

If self-acceptance and prejudice were uncorrelated, one would have no justification for

TABLE 1
OCCUPATIONAL GROUPS REPRESENTED
IN THE OSGOOD HILL SAMPLE

Males ($N = 20$)		Females ($N = 30$)	
Clerics	4	Teachers (nursing school)	5
Graduate students	3	Nuns	2
Social workers	3	Nursing students	6
Dentist	1	Nurses/nursing	12
Teachers	3	supervisors	
Policemen	2	Graduate students	2
Government	2	Social workers	3
Industry	2		
Total	20	Total	30

hypothesizing that a change in self-acceptance would be associated with a change in prejudice. On the other hand, if these two variables were perfectly correlated, the hypothesis would be tautological. The initial hypothesis to be tested, therefore, is that the higher one's level of self-acceptance (*ETA*), the lower one's level of prejudice (*PR*). This degree of association, as measured by the Pearson product-moment correlation coefficient is $r = -.32$ ($N = 41$, $p < .05$, one-tailed). In other words, the degree of association between *ETA* and *PR* is significantly different from zero and in the predicted direction.

Control Group—Comparability and Stability

Table 2 contains the test-retest scores for the control group and the before-after scores for the experimental group. A series of *t* tests were performed in order to determine empirically the degree of similarity between experimentals and controls. None of the resulting *t*'s reached significance, with most *p*'s being greater than .50. On the basis of these results, it is assumed that the members of the control group represent a population comparable to the experimentals on the major variables.

TABLE 2
BEFORE-AFTER SCORES FOR CONTROL GROUP AND
BEFORE SCORES FOR EXPERIMENTAL GROUP

	Control group ($N = 11$)		Experimental group ($N = 30$)	
	Before	After	Before	After
Number of ego-threatening statements ($\#ET$)	11.0	12.0	13.5	13.2
Self-Acceptance (<i>ETA</i>)	66.0	65.0	55.0	67.0
Prejudice (<i>PR</i>)	47.5	47.5	46.2	42.0

It can be seen from Table 2 that among the members of the control group: $\#ET$ increased slightly, and ETA and PR both decreased slightly. Using a t test for dependent samples (Blalock, 1960), it was observed that none of the resulting t 's reached the .60 level of significance. On the basis of these results, it is assumed that any change found among experimentals cannot be attributed to the main effects of instrument instability and/or practice.

Effects of Sensitivity Training

It was hypothesized that $\#ET$ would not change as a result of sensitivity training. Examination of Table 2 reveals that $\#ET$ decreased slightly over this 2 week period. Using a t test for dependent samples, it was found that for $\Delta\#ET$ (change in $\#ET$), $t = .70$ with an associated $p < .45$, two-tailed ($N = 30$). We can therefore assume that sensitivity training had no appreciable effect upon the absolute number of ego threatening statements generated by an individual on the Sentence Completion Test.

The next prediction was that, as a result of sensitivity training, self-acceptance (ETA) would increase, and prejudice (PR) would decrease. ETA went from 55.0% to 67.0%, while PR went from 46.2 to 42.0. The differences between these means (t test for dependent samples) are significant. For ETA , $t = 2.58$ ($N = 30$, $p < .01$), while for PR , $t = 2.54$ ($N = 30$, $p < .01$).

Thus far it has been shown that the experimentals change significantly while the controls do not. The critical test, however, is whether the experimentals change more than the controls. A Mann-Whitney U test (Siegel, 1956) was performed on the "difference between the differences" within experimentals and controls. This analysis yielded a $z = 1.65$ for ETA ($N_1 = 11$, $N_2 = 30$, $p < .05$, one-tailed) and a $z = 1.76$ for PR ($N_1 = 11$, $N_2 = 30$, $p < .04$, one-tailed). In other words, not only do the experimentals change while the controls do not, but also the experimentals change significantly more than the controls.

Change in ETA (ΔETA) versus Change in PR (ΔPR)¹¹

The main thrust of this investigation concerns the relationship between changes in

¹¹ For the purpose of this and the following analyses, the 8 of 11 control-group members who returned us-

self-acceptance and changes in prejudice. Splitting the sample at the median ΔET (+8%, i.e., 8% increase self-acceptance) ΔPR (-2.0, i.e., 2-point decrease in prejudice level) yields the following 2×2 contingency table:

	High ΔETA	Low ΔETA
High ΔPR	15	8
Low ΔPR	4	11

A test of the significance of this distribution yields a $\chi^2 = 4.0$ which is significant at the .02 level, one-tailed ($df = 1$).¹²

In a further analysis of the data, individual change scores on ETA were examined more closely. There appeared to be a sharp discontinuity in the distribution of ΔETA scores. Several people increased a moderate amount in ETA (8-14%), but then the next highest change was 21%. There were 13 people who increased 21% or more in self-acceptance. When we examine this group of high $+\Delta ET$'s versus the remainder of the sample, the following results emerge. The high $+\Delta ETA$ group decreases an average of 8.0 points on PR ($N = 13$, $t = 3.0$, $p < .01$, one-tailed), while the remainder of the sample decreases an average of 2.0 points on PR ($N = 25$, $t = 1.3$, $p < .12$, one-tailed). A Mann-Whitney U test on the difference between these differences yields a $z = 1.76$ ($N_1 = 13$, $N_2 = 25$, $p < .04$, one-tailed). In other words, those who increase a great deal in self-acceptance ($\Delta ETA > 20\%$) will decrease significantly more in prejudice than those who decrease in self-acceptance or increase only a moderate amount.

Alternative Hypotheses

Several alternative explanations for the observed changes in ETA and PR were examined. In order to check for the possible contaminating influence of the regression effect, initial ETA and PR scores were correlated with changes in ETA and PR . For ETA , $r = -.21$ which is not significantly different from zero. For PR , $r = -.20$ which is in a direction opposite to that expected if regression were

able responses after the lab were added to the 30 experimentals. These 8 people changed as much (percentage-wise) in ETA and PR after the laboratory as did the experimentals. In addition, like the experimentals, they did not change in $\#ET$. This raises our available population from $N = 30$ to $N = 38$.

¹² A test of the causal linkage, discussed elsewhere (Rubin, 1966, 1967), suggests that a significant amount of the observed reduction in prejudice results from increases in self-acceptance.

operating, but is also not significantly different from zero. Statistical regression is therefore ruled out as a possible alternative hypothesis.¹²

There were five separate training groups and six individual trainers (one group had two trainers) which made up the Osgood Hill sensitivity training program. Two of the trainers were females, one of them a Negro. It is therefore possible that changes in *ETA* and *PR* might somehow be related to individual trainer styles or some other factors unique to a particular training group. A Kruskal-Wallis one-way analysis of variance (Siegel, 1956) between the five groups on all major variables was performed, and none of the resulting *H*s reached the .50 level of significance, two-tailed. From this it can be assumed that there was no significant trainer effect nor can the observed changes be attributed to some other factor unique to any one of the training groups. On the other hand, the fact that there was at least one Negro in each group makes it impossible to assess the impact of interracial contact on the participants' attitudes.

DISCUSSION

One of the central hypotheses in this study was that changes in self-acceptance would be associated with changes in prejudice. The hypothesis in its original form was confirmed. A further analysis of the data suggested a threshold effect in the relationship between increases in self-acceptance and decreases in prejudice. Perhaps, where sensitivity training really took (in the sense of great increases in self-acceptance), those involved may have been better able to immediately make the mental transfer from self-acceptance to decreased prejudice. The others may have needed some period of incubation in order for this transfer to occur.

Support for this interpretation is provided by Katz, Sarnoff, and McClintock (1956, 1957) who found that the changes in prejudice immediately after a self-insight manipulation, though significant, were no greater than the changes produced through providing more information about minority groups. Five weeks later, however, the significant effects of reduced prejudice were still found in the group receiving the self-insight materials, whereas in

the group receiving the information the effects were no longer significantly different from the control group. In other words, more stable

changes were observed in the group receiving the self-insight manipulation. The written case study utilized by Katz et al. to increase self-insight is certainly less intensive than a 2-week sensitivity training laboratory and may well be less powerful. It is possible therefore that decreases in prejudice will persist after the laboratory and, in fact, may become more marked in the group which experienced only moderate increases in self-acceptance. This is a hypothesis which could not be tested due to the fact that it was necessary to provide a full feedback session for the laboratory participants prior to their departure.

Another issue raised in this study concerns the interaction between theory and research. The independent variable—sensitivity training—consists of many forms of learning. Even within the training group—the core element of the total program—many phenomena occur over a 2-week period (e.g., feedback upon one's effect on other people). Ideally, in any change study, the researcher would like to have additional experimental controls to be able to compare the effects of sensitivity training versus some other change strategy, or to control for various elements within the training group. When these are unavailable, the researcher must rely on his theory to compensate for this lack of additional controls. There are two instances in the present study where this interaction is critical if the observed changes are to be attributed to sensitivity training.

Within the design utilized, it was possible only to compare changes over a period of ordinary activity with changes assumed to result from sensitivity training, making a Hawthorne effect a possibility. There is evidence, however, to suggest that it is not just an unusual experience which is generating the results. The fact that changes were observed only on those variables hypothesized to change and no changes were observed on those hypothesized to remain stable tends to support this statement. If a Hawthorne effect were operating, the variables influenced should be random rather than following the theoretically predictable pattern observed.¹⁴

¹²In addition, the sample was split into thirds according to initial *ETA* and *PR* scores, and changes within each group were examined. Each group changed in the predicted direction. This also suggests that even those initially low in self-acceptance (0-40%) can be "reached" by sensitivity training.

¹⁴In the experiment from which this paper is drawn (Rubin, 1966), other variables not discussed here were shown, as hypothesized, to remain stable after sensitivity training. One such variable was psychological anonymity which was shown to condition the influences of sensitivity training—the lower an individual's level of anonymity, the more he increased in

Second, this study rests upon a theoretical framework which lays out a very specific sequential chain of events. The tradeoff, generally speaking, is that if such a chain breaks down anywhere, little is known about the hypothesized effects of the independent variable. If, on the other hand, the chain of theory does hold up, then the probability has increased considerably that it is the independent variable (in this study, sensitivity training) which is causing the observed changes.

The experimental subjects studied in this research were not a representative sample of the general population. Aside from their atypical educational and occupational levels, they also shared a certain level of "motivation to attend a laboratory." It is not yet known what personality variables, for example, differentiate those who are "motivated to attend" from those who have no such motivation. Even if knowledge of these parameters did exist, it would then have to be demonstrated that they have relevance in terms of differential learnings resulting from training. Further research is needed to answer these questions before the findings of this study can be widely generalized.

Finally, the reader has undoubtedly noticed that, by changing a few words (e.g., training group to therapy group, trainer to therapist), this study could have been concerned with the effect of client-centered psychotherapy upon prejudiced attitudes. Each provides the elements of psychological safety, support, and opportunities for reality testing assumed necessary to affect an increase in an individual's level of self-acceptance and consequently, by our model, decrease his level of ethnic prejudice. To the extent that future research and practical experience substantiate the conclusions drawn from the present study, a step has been taken toward solving a problem posed by Adorno et al. (1950) some 15 years ago:

Although it cannot be claimed that psychological insight (self-insight) is any guarantee of insight into society, there is ample evidence that people who have the greatest difficulty in facing themselves are the least able to see the way the world is made. Resistance to self-insights and resistance to social facts are contrived, most essentially, of the same stuff. It is here that psychology may play its most important role. Tech-

self-acceptance and decreased in prejudice. Anomy, furthermore, is positively related to an individual's tendency to acquiesce (Rubin, 1966, 1967). These data are presented here to point out that if acquiescence were a major factor causing change, the results would be opposite to those observed.

niques for overcoming resistance, developed mainly in the field of individual psychotherapy, can be improved and adapted for use with groups and even for use on a mass scale [p. 976].

REFERENCES

- ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper & Row, 1950.
- BLALOCK, H. M. *Social statistics*. New York: McGraw-Hill, 1960.
- BUNKER, D. Individual applications of laboratory training. *Journal of Applied Behavioral Science*, 1968, 1, 131-147.
- DORRIS, R. J., LEVINSON, D., & HANFMANN, E. Authoritarian personality studied by a new variation of the sentence completion technique. *Journal of Abnormal and Social Psychology*, 1954, 49, 99-108.
- GORDON, T. What is gained by group participation. *Educational Leadership*, January 1950, 220-226.
- HARDING, J., & SCHUMAN, H. An approach to the definition and measurement of prejudice. Unpublished manuscript, Harvard University, 1961.
- JOHNSON, P. C. *Statistical methods in research*. Englewood Cliffs, N. J.: Prentice-Hall, 1949.
- JONES, R. M. *An application of psychoanalysis to education*. Springfield, Ill.: Charles C Thomas, 1960.
- KATZ, D., SARNOFF, I., & MCCLINTOCK, C. M. Ego defense and attitude change. *Human Relations*, 1956, 9, 27-45.
- KATZ, D., SARNOFF, I., & MCCLINTOCK, C. M. The measurement of ego defense as related to attitude change. *Journal of Personality*, 1957, 25, 465-474.
- ROGERS, C. *Client-centered therapy*. Boston: Houghton Mifflin, 1951.
- RUBIN, I. M. Sensitivity training as a means of reducing prejudice. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Alfred P. Sloan School of Management, 1966.
- RUBIN, I. M. The reduction of prejudice through laboratory training. *Journal of Applied Behavioral Science*, 1967, in press.
- SCHEIN, E. H., & BENNIS, W. G. *Personal and organizational change through group methods: The laboratory approach*. New York: Wiley, 1965.
- SCHUMAN, H., & HARDING, I. Sympathetic identification with the underdog. *Public Opinion Quarterly*, 1963, 27, 230-241.
- SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.
- WESCHLER, I. The self in process: A sensitivity training emphasis. In I. Weschler & E. H. Schein (Eds.), *Issues in human relations training*. Washington, D. C.: National Education Association, National Training Laboratories, 1962. (NEA NTL Selected Readings Ser. No. 5)
- WYLIE, R. *The self concept*. Lincoln: University of Nebraska Press, 1961.

(Received March 16, 1966)

COMPLIANCE, CONSISTENT CONFORMITY, AND PERSONALITY¹

MARJORIE H. KLEIN

Wisconsin Psychiatric Institute

Various conformity and nonconformity responses were evoked by a complex influence technique in which sources, arguments, and measurement settings were varied. Personality correlates of compliance (public without private conformity to authorities) were investigated. When compared with Ss responsive to different influence pressures, compliant Ss were found to share a limited approval orientation. When contrasted with groups showing different responses to the same source, compliant Ss have only superficial approval needs, avoid emotional involvement, prefer intellectual defenses, and are pragmatic, secure, and autonomous. Ss who conform consistently (in public and in private) to the same authorities share the superficial approval orientation, but also have more general approval needs and lower self-esteem and prefer regressive defenses.

In studies of social influence, correlations have emerged for varieties of conformity behavior and some personality factors, especially self-esteem, defensiveness, and needs for approval, external direction, or structure. These relationships have varied in strength, however, and have not been consistently replicated. It is increasingly clear that conformity is a complex phenomenon encompassing a range of responses with many different functions. Strong and predictable conformity-personality relationships cannot be expected to emerge unless the conformity process involved is clearly defined, and situational factors (e.g., content, source, source-subject relationship, demand character of influence and measurement settings) are taken into consideration.

This study follows Kelman's (1961) model for social influence in which three processes (compliance, identification, internalization) are differentiated. Compliance is defined as a limited, superficial public conformity, unaccompanied by private attitude change. The response, defined in terms of external demands, is oriented to the attainment of approval or some specific reward under control of the influence agent. Performance is limited to the agent's surveillance. Identification and internalization are more private processes. Identification is oriented to the establishment of a self-defining relationship with an

attractive role model or partner. Response performance is limited to situations of role salience. In internalization, the induced response is integrated with attitudes or values that the individual already possesses. While the source credibility is important initially, the response becomes autonomous, sustained by the appropriateness of the content to the value context.

While this model gives each process a distinct motivational or functional orientation, it allows the balance of internal (personality) and external (situational) processes in activating or sustaining the motivation to vary. Any type of conformity can be closely and/or generally tied to personality, but it is also possible that personality factors may be quite limited in scope, or that situational pressures alone are more important.

This paper reports an examination of the compliance response in order (a) to test the hypothesis that compliance is associated with the need for social approval, and (b) to explore the generality and functional significance of the approval needs involved. Methodological features are used to isolate and clarify the personality-compliance relationships: influence pressures are noncoercive; simultaneous pressures for each of the three processes are offered, and subjects are free to respond selectively or to show a mixed conformity pattern. Only subjects showing preferential sensitivity to compliance are considered. Comparison groups include (a) subjects responsive to pressures for identification or internalization as well as those showing no change, and (b) those giving other responses to the same authoritative source (consistent conformity, no change, negative change). Personality measures included tap both general and specific approval needs, self-esteem, interpersonal perceptions, and attitudes

¹ This article is based on a dissertation submitted in partial fulfillment of the requirements for the PhD in social psychology at Harvard University. The author gratefully acknowledges the invaluable assistance of Herbert C. Kelman who served as major advisor, and the thoughtful criticism of David Marlowe. The research itself was carried out at the University of Wisconsin with the generous support of the Department of Psychology and the Wisconsin Psychiatric Institute.

toward interpersonal conflict, influence pressures, and personal change.

METHOD

Influence Procedure

The influence technique administered to define conformity groups was given over three sessions to undergraduates in introductory psychology ($N=210$).

Session 1 (pretest). The target of the influence procedure was the subject's description, on an adjective-rating scale, of his goals for personal development over the college years. This, the Personality Goals Test (PGT), containing 100 items rated on a 7-point scale of desirability, was pretested in introductory psychology quiz-section meetings. The regular section leaders introduced it as part of a university-wide survey.

Session 2 (influence communications and public retest). Three weeks later, the following two-part task was given in the quiz sections by the section leaders.

1. *Influence Communications.* Influence communications were introduced as the PGT survey report. Separate paragraphs described three personality-trait patterns, each containing 15 fairly cohesive and distinct PGT adjective clusters, attributed to three sources, university administrators, most university students, and social philosophers. Sources were presented in standard order. Adjectives were varied systematically across thirds of the sample (ABC, CAB, BAC) to balance specific communication content. (a) The *compliance* paragraph (authoritative source-approval orientation) presented adjectives supposedly considered desirable by the university administration for graduating college students. Supporting argument stressed approval rewards. (b) The *identification* (attractive source-role orientation) paragraph presented adjectives supposedly rated desirable as goals by the average university student and by student leaders, with campus adjustment and peer relationships stressed as rewards. (c) The *internalization* communication (credible source-value orientation) presented adjectives supposedly considered desirable by social philosophers in enabling the young adult to contribute to society and implement personal values.

Subjects were instructed to read the reports carefully and to be prepared to answer questions about the content.

2. *Public Retest.* To create a measurement condition appropriate for compliance (surveillance by authority figures), the subjects had to be convinced that PGT retest responses would be seen by the university administration. Immediately after reading the influence communications, the subjects were asked to retake the PGT under administration auspices. The reason given was that administration concern about the wide discrepancy between administration and student viewpoints had led them to request trials with an alternate PGT form to test the validity of the rating procedure. The subjects were told that

the university administration would process the protocols, and were asked to provide their full names and identification numbers. The PGT form used was similar to the pretest form with only minor variations in details of the rating procedure.

Session 3 (private and other retests). In order to create measurement conditions appropriate for identification and internalization (nonsurveillance, role salience and value relevance), it was necessary to readminister the PGT in a relatively private or anonymous setting where the university administration would not be expected to have access to results. This was done by administering a superficially altered format of the PGT, in a different setting and research context. The investigator (who had not previously appeared before the students) was introduced at the subjects' formal lecture meeting as a graduate student from another university conducting a study of the college years. The PGT was included as part of the initial screening, and the subjects were assured of anonymity and of the confidentiality of results.

Additional retests were designed to distinguish identification and internalization groups among those showing private conformity. For identification, it was important to establish, as specified in Kelman's model, that the response formed the requirements for the student role. For internalization, the response had to be clearly anchored within the value context. Subjects were asked to indicate, on a separate check list, which of the 45 critical adjectives they would endorse in each of the following conditions: (a) counseling a student regarding adjustment to student life (role salience), (b) giving a speech about qualities necessary for a young adult to make a contribution to society (value context), and (c) by a university administrator describing goals for college students (to assess retention of the compliance response).

Scoring Criteria for Influence Processes

Pretest and retest protocols were scored for each conformity response according to the definitions below. Cutoff points for change were determined by examining the distribution of change scores for each adjective cluster separately, and by taking the upper, middle, and lower thirds ($N=210$) to represent high change, no change, and negative change, respectively.

Compliance. High change toward the administration communication from pretest to public retest (high public change), and high movement away from the administration communication from public to private retest (high private drop).

Identification. High movement toward the average university student communication from pretest to private retest (high private change), and high endorsement of the same adjectives for the student role.

Internalization. High movement toward social philosopher communication from pretest to private retest, and high endorsement of the same adjectives in the value context.

Composition of Groups for Analyses

Part 1. Analyses reported in Part 1 contrast the compliance group with a composite control group, defined as follows:

1. Compliance A. Subjects high in compliance but low or moderate in both identification and internalization ($N = 16-10$ females, 6 males). This is referred to as Compliance A in order to distinguish it from the alternative Compliance B group in Part 2.

2. Composite Control Group. This includes three subgroups: (a) subjects high in identification but low or moderate in both compliance and internalization ($N = 21-13$ females, 8 males), (b) subjects high in internalization but low or moderate in both compliance and identification ($N = 15-10$ females, 5 males), and (c) no change group—subjects with moderate change scores (i.e., neither positive nor negative change) for compliance, identification, and internalization ($N = 14-9$ females, 5 males).

Part 2. Analyses reported in Part 2 contrast four groups, defined by their response to the administration communication only²:

1. Compliance B. Subjects showing high public change and high private drop ($N = 17-13$ females, 4 males).

2. Consistent Conformity. Subjects high in public and private change ($N = 18-15$ females, 3 males).

3. No Change. Subjects with moderate change (i.e., no change) in both public and private retests ($N = 23-12$ females, 11 males).

4. Negative Change. Subjects showing negative change in both public and private retests ($N = 25-13$ females, 12 males).

Personality Testing

About 1 month after the completion of the influence series, subjects meeting criteria for the various groups were invited by letter to participate in the personality testing. The subjects were told that they were taking part in a comprehensive study of college students, reminded that the investigator was a graduate student from another university, and assured that test responses would be kept confidential with names replaced by code numbers before scoring. The following tests were administered in three group sessions, according to the standard instructions for each instrument: F scale (Couch & Keniston, 1960), Antipsychological Mindedness Scale (Sharaf, 1960), Marlowe-Crowne Social Desirability (M-C SD) scale (Crowne & Marlowe, 1964), Role Construct Repertory Test (RCRT—Kelly, 1955); stimulus pic-

tures for need affiliation and fear of rejection, (Atkinson, 1958); Cohen Self-Esteem and Defense Preference Scale, consisting of an actual and ideal form in which the subject indicates actual and ideal solutions to interpersonal-conflict situations involving different needs.³ Last presented were open-ended and objective questions specially designed to tap self-described goals, perceptions of influential attributes of others, and assessments of the nature and causes of personal change in actual and hypothetical situations.

Test Scoring

All tests were scored according to the standard procedures. The self-esteem score was determined by the amount of correspondence between actual and ideal self-descriptions. Defense-preference scores derived from this test consist of the frequency with which problem solutions fall into five defense categories: avoidance, reaction-formation, regression, intellectualization, and projection. Need-affiliation and fear of rejection scores were obtained by breaking affiliation imagery into positive and negative categories, according to the procedure suggested in Atkinson (1958).

RCRT responses, consisting of the constructs used to differentiate significant others, and all responses to open-ended and objective questions about hypothetical change situations were coded by the experimenter according to categories derived from Kelman's model.

Approval orientation (compliance). Stress on gaining approval or status; perception of others' power, status, or approval-giving capacities; change seen as behavioral conformity to external demands.

Affiliation or identity orientation (identification). Stress on identity, role enactment and role relationships, or degree of similarity to others; perception of others' attractiveness as role models or relationship objects; change seen as change in identity or role behavior.

Value or cognitive orientation (internalization). Stress on values, ideals, attitudes, ability to gain cognitive clarity, or internal consistency; perception of the credibility, values, or value consistency of others; change seen in terms of values, ideals, attitudes, or means-end relationships.

RESULTS

Compliance A and Composite Control Groups

In order to test the hypothesis that compliance is related to the need for social approval, Compliance A subjects were contrasted with the composite comparison group. One-tailed t tests were used for hypothesized relationships; two-tailed tests, for exploratory analyses.

Of the six approval measures tested, only two, the F scale, for males, and the approval code for

² As these groups were formed only on the basis of responses to the compliance induction, they are not identical in composition to the groups based on the more complex criteria. There is some overlap between Compliance A and Compliance B groups, but new subjects are included in the B group who showed a mixed conformity response (e.g., high both in compliance and in identification or internalization) when all inductions were considered.

³ A. R. Cohen, "A Report on Some Explorations of Self-Esteem," unpublished report, 1954.

open-ended questions, for females, showed higher scores for the compliance group (see Table 1). Also, there was a significant reversal in approval perceptions for the RCRT, for females. Apparently the heightened approval orientation of compliance-group subjects is limited in scope to measures dealing directly with attitudes toward authority figures, social approval, and conventional, approval-seeking behavior. These differences do not hold for more generalized perceptions and needs (e.g., as in the RCRT, where freely chosen significant others are evaluated), suggesting that the approval needs shared by compliant subjects are relatively overt and superficial, focused primarily on attitudes that make conformity to conventional standards and demands of authority figures desirable and compatible.

Exploratory analyses reported in Table 2 show the compliance group also to have stronger anti-intracceptive attitudes, and compliance-group males to have a greater preference for the defense of intellectualization. Apparently the compliant subject is one who actually prefers to limit his emotional involvement in social conflict or responds by making an intellectual, surface adaptation.

Finally, results for two of the open-ended questions, dealing specifically with sources of self-satisfaction, suggest additional goals that may be associated with compliance. All subjects made more frequent mention of achievement as a source of self-satisfaction than control subjects, and males stressed emotional support obtained

TABLE 1
APPROVAL NEED SCORES FOR COMPLIANCE A
AND COMPOSITE CONTROL GROUPS

Approval measure		Compliance	Composite control	<i>p</i>
Open-ended questions	Male	2.83	2.72	.13
	Female	3.70	2.03	2.52**
Objective questions	Male	1.50	1.94	.64
	Female	1.90	2.28	.77
F scale	Male	5.34	4.46	1.83**
	Female	4.82	4.64	.50
RCRT	Male	1.50	1.50	.00
	Female	.90	2.31	2.38**
M-C SD scale	Male	12.67	14.11	.50
	Female	14.40	15.12	.34
TAT-fear of rejection	Male	.33	.56	.64
	Female	.70	.62	.25

* One-tailed *p* values except for RCRT.
** *p* < .05 level.

TABLE 2
TEST SCORES FOR COMPLIANCE A AND
COMPOSITE CONTROL GROUPS

Measure		Compliance	Composite control	<i>p</i>
Anti-intracception scale	Male	4.55	4.11	1.09
	Female	4.10	3.62	1.85*
	Both sexes	4.27	3.80	2.08**
Intellectualization defense	Male	3.00	1.50	2.40**
	Female	2.30	2.38	.18
Satisfaction with achievement	Male	.83	.33	1.67
	Female	.40	.18	1.38
	Both sexes	.56	.24	2.18*
Satisfaction with emotional support	Male	.50	.11	2.14*
	Female	.50	.44	.31

* Two-tailed *p* values; row for both sexes omitted when non-significant.

* *p* < .10 level.

** *p* < .05 level.

from personal relationships. These responses may simply reflect the more conventional outlook of the compliance group, but it is also possible that such goals are ancillary aims or important by-products of agreeable, compliant behavior.

Compliance B and Consistent Conformity

Table 3 presents results of Scheffé analyses of variance of test scores for four groups, defined according to responses to the authoritative source only. Results for some approval measures and for all other measures yielding *F* ratios for group effects significant at the 5% level or beyond are tabled, together with the outcome of Duncan multiple-range tests for between-group differences.⁴

The most consistent pattern of results applies to the two conformity groups, compliance (public but not private conformity) and consistent conformity (both public and private conformity).

⁴ These analyses also include the two nonconformity groups (no change and negative change). Some differences for these groups parallel those for the two conformity groups. The no change group is generally close to the compliance group; negative changers and consistently conforming subjects have features in common.

Sex was also included as a factor in these analyses, but only significant group effects are reported. There were only a few significant Sex × Group interactions, and they do not fall into a consistent pattern. Analyses of variance carried out separately for males and females support the overall results, but effects are stronger for males than for females.

TABLE 3
TEST SCORES FOR COMPLIANCE B, CONSISTENT CONFORMITY,
NEGATIVE CHANGE, AND NO CHANGE GROUPS

Measure	Compliance C	Consistent conformity Y	Negative change R	No change N	Duncan multiple- range test*
F scale	4.90	4.74	4.45	4.63	
Open-ended questions—approval code	2.47	3.00	2.56	3.00	
M-C SD scale	12.53	15.22	16.32	13.65	RY NC**
Satisfaction with own values	.71	.11	.28	.30	CNRY***
Satisfaction with approval	.18	.83	.32	.61	YNRC**
Objective questions—approval	2.24	1.33	1.92	2.21	CNRY**
Self-esteem scale	13.00	8.94	10.20	11.30	CNRY***
Intellectualization defense	2.76	2.44	1.48	1.91	CYNR***
Regression defense	.59	1.44	1.64	.83	RY NC***

* Groups not joined by underlining differ at the 5% level of significance or less for Duncan multiple-range test.

** $p < .05$ (for F ratio).

*** $p < .01$ (for F ratio).

While it is noteworthy that they fail to differ in overt approval attitudes (F scale, approval code for open-ended questions), the consistently conforming subjects gave the more socially desirable M-C SD scale responses, indicative of more generalized needs for approval. Also, the two groups differed in factors reported as sources of self-esteem on one of the open-ended questions. Consistently conforming subjects especially valued others' approval reactions; compliance-group members stressed personal values and standards. Finally, there was an important difference in responses to objective questions. Those in the compliance group more frequently interpreted others' conformity as openly motivated by desires to manipulate or win approval from the influence agent.

As Table 3 also indicates, compliant subjects scored higher in self-esteem than consistently conforming subjects and differed in defense preference. Both groups showed a relatively strong

inclination for intellectualization, but this defense was especially preferred by the compliance group, while the consistent conformity group also stressed regression.

Taken as a whole, these results suggest that the approval needs of the compliance group were not only superficial and limited in scope, but also were instrumentally anchored and superimposed on a relatively autonomous, rational personality structure in which internalized standards are at least as important as external approval for self-esteem. In contrast, approval needs were more generalized, and social reinforcement seemed more important for ego support in the less secure, more dependent, and regressive consistent conformity group.

Generality of Compliance and Consistent Conformity Responses

In interpreting differences between compliance and consistent conformity groups, it is important to know whether the response patterns used to define these groups are specific to the authoritative source or reflect more general differences in the stability of conformity behavior. To consider this question, the groups' public change and private drop scores for student and social philosopher communications were compared. As Table 4 shows, no differences comparable to those for the administration communication were found, suggesting that personality test results can be interpreted in terms of specific differences in responsiveness to the authoritative source.

DISCUSSION

Because the number of significant differences relative to the number of analyses attempted is

TABLE 4

PUBLIC CHANGE AND PRIVATE DROP SCORES OF COMPLIANCE AND CONSISTENT CONFORMITY GROUPS FOR ADMINISTRATION, STUDENT, AND PHILOSOPHER COMMUNICATIONS

	Compliance	Consistent conformity	t ($df = 33$)
Administration			
Public change	13.12	15.17	not relevant
Private drop	11.47	4.89	
Student			
Public change	5.24	3.61	<1.00
Private drop	4.53	1.61	1.30
Philosopher			
Public change	3.88	6.67	<1.00
Private drop	2.41	1.17	1.11

small because the number of subjects in each group is modest, and not all differences hold for both sexes; the results must be considered only as tentative. The pattern that does emerge suggests divergent functions for compliance and consistent conformity, and different levels of generality for some of the personality-conformity relationships.

Compliance and consistent conformity groups seem to have a common approval orientation, sharing just those surface attitudes necessary to support common elements in their conformity behavior. These attitudes, however, relate to needs and trait patterns that suggest different interpretations for the two processes.

In compliance, the approval orientation is limited to the superficial attitudinal level. Related personality factors suggest that the response may function primarily as a rational, even calculated process of accommodation in which external demands are met or social conflict avoided, without deep emotional involvement or personal commitment. Indeed, compliance may function in the long run as a form of conformity avoidance, an expedient compromise between adjustment and autonomy needs most readily available to the practical, rational, and somewhat independent person. In support of this view of compliance as serving independence needs, it is perhaps meaningful to note that the compliance and no change groups were similar in some respects.

Factors associated with consistent conformity point to a more familiar interpretation of this response as an ego supportive or defensive process. Approval needs are more general and are also associated with insecurity and dependency. In this context, the more generalized conformity behavior may function to provide structure where internal standards are inadequate, and insure feedback to enhance self-esteem in the absence of personal security. In Kelman's terms, it seems reasonable to view this response as a type of classical identification (parent-child) whereby the consistently conforming subject meets role demands of university authorities, not just to placate them but also to be able to define himself as acceptable in their eyes.

While it may seem somewhat unusual to find conformity, even of the compliance type, associated with independence, rationality, and security, this relationship may reflect the special social developmental bind of the college student. Assuming most college students encounter explicit demands for conformity to institutional rules and implicit pressures to develop adult autonomy, a response such as compliance that reconciles short-term dependency needs with long-term autonomy goals seems the more rational, mature adaptation

to influence pressures emanating from university authorities. In contrast, a generalized conformity one that goes beyond merely to situation demands, seems the more infantile solution.

This view of the differences between compliance and consistent conformity has been based in part on the assumption that the M-C SD taps differences in the absolute level of the subject's general needs for approval. When, however, the focus is shifted somewhat, a more parsimonious interpretation can be developed.

If we compare features of the conformity and personality measurement sessions for this study, it is clear that personality testing sessions were closer in atmosphere to the private rather than to the public authoritative conformity measurement session. Confidentiality was assured; the investigator was explicitly disassociated from the university administration. Accordingly, one might view the M-C SD-scale responses of the two conformity groups as being directly analogous to their private conformity retest behavior; that is as reflecting differences in their self-presentation needs in a relaxed or private test setting.

From this assumption that M-C SD scale and conformity responses may be similarly influenced by situational factors in the test setting, it follows that the compliance and consistent conformity groups would obtain equally high M-C SD scores if tested in public authoritative circumstances more similar to the public conformity measurement setting. In this case, the results of this study could be reconciled with research where personality and conformity tests were administered at the same time and public conformity has been associated with high M-C SD-scale scores (Crowne & Marlowe, 1964).

This situationally oriented approach to conformity-personality relationships has compelling implications for the methodology of conformity research. When situational factors are deliberately manipulated as they are in social influence studies, both conformity and personality measures may become unusually vulnerable to situational variance. Also, superficial personality factors shared by subjects showing general similarities in conformity may obscure equally important functional distinctions that are more variable, idiosyncratic, and certainly more difficult to isolate experimentally.

REFERENCES

- ATKINSON, J. W. (Ed.) *Motives in fantasy, action and society*. Princeton, N. J.: Van Nostrand, 1958.
 COUCH, A., & KENISTON, K. Yeasayers and naysayers: Agreeing response set as a personality variable.

- Journal of Abnormal and Social Psychology*, 1960, 60, 151-174.
- CHOWNE, D., & MARLOWE, D. *The approval motive*. New York: Wiley, 1964.
- KATZ, G. A. *The psychology of personal constructs*. New York: Norton, 1955.
- KELMAN, H. C. Processes of opinion change. *Public Opinion Quarterly*, 1961, 23, 57-78.
- STANLEY, M. R. An experiment in the theory and measurement of identification. Unpublished doctoral dissertation, Harvard University, 1960.

(Received March 14, 1966)

Journal of Personality and Social Psychology
1967, Vol. 5, No. 2, 245-249

NEED ACHIEVEMENT AS A FUNCTION OF THE RACE AND SEX OF FIGURES OF SELECTED TAT CARDS

GLORIA COWAN¹

AND

FAVE J. GOLDBERG

Wayne State University

Morehouse College

In a study of achievement motivation in Negro female college students, the effects of characteristics of the figures on selected TAT cards and on need for Achievement were determined by independently varying sex and race of the stimulus figures. Ss were divided into 4 groups, each group was given 1 of the 4 following types of pictures: Negro female, Negro male, white female, white male. Results indicated that for *n* Achievement scores uncorrected for protocol length, both race and sex of the figures were significant. A measure of productivity, word count, was affected by race of picture but not by sex. For *n* Achievement scores corrected for length of protocol, only the sex variable was significant; i.e., more *n* Achievement motivation was expressed by Negro females to male figures, irrespective of race, than to female figures. The highest scores were obtained in response to pictures of Negro males for both productivity and *n* Achievement measures, corrected and uncorrected.

The problem of defining the cue characteristics of pictures used in thematic apperception measurement of the achievement motive (*n* Achievement) is particularly crucial in attempting to assess such motives in Negro subjects. Studies comparing Negroes and whites have been reported (Merbaum, 1962; Mingione, 1965; Rosen, 1959), and studies comparing northern and southern Negroes have also been reported (Nuttall, 1964). Very little has been published on the effects of race and sex on Negro subjects, particularly female subjects. Since *n* Achievement is functionally related to level of aspiration and risk taking (Atkinson, 1958) and other socially important behaviors, it is of value to reexamine factors affecting assessment of *n* Achievement in Negroes. It is possible to raise the general question posed by the use of such projective devices as the TAT; namely, is material likely to be more pertinent to the subject's needs, motives, etc., when he is able to "identify" with the central figure, especially with regard to race and sex?

In the late 1940's and early 1950's, a few investigators were concerned with the effect of

race of the figures on TAT scores (not *n* Achievement) in studies using Negro subjects, that is, whether Negro figures would facilitate greater identification in Negro subjects than would white figures. Thompson (1949) developed the T-TAT which included changed racial characteristics on 10 of the Murray cards. Although Thompson (1949) found significant increases in the number of words and parts of speech for Negro male subjects when using Negro figures as opposed to white figures, research following the development of the T-TAT (Cook, 1953; Korchin, Mitchell, & Meltzoff, 1950; Riess, Schwartz, & Cottingham, 1950; Schwartz, Riess, & Cottingham, 1951) indicated that race of the figures did not affect productivity or "idea count." Since contradictory results were obtained by Thompson and those who attempted to follow up his findings, no conclusion can be reached on the basis of these studies. Two criticisms of Thompson's work are notable: the design was such that the white and Negro figures and situations were not comparable; the figures were so changed from the original TAT, in expression and situation, as to render the tests dissimilar.

The other line of research which independently converges on the present study is the work done

¹ This study was carried out while the first author was at Spelman College.

on the effect of sex of the figures on n Achievement in white females. Veroff, Wilcox, and Atkinson (1958) found that University of Michigan female college students obtained higher n Achievement scores in response to male characters than to female characters. Fields (in McClelland, Atkinson, Clark, & Lowell, 1953) found that achievement scores at University of Maryland women increased under arousal conditions when the arousal was operated in terms of social acceptance, not under intellectual arousal. French and Lesser (1954) found that motivation scores of female white college students at Hunter College in New York were higher when male figures were used under intellectual arousal conditions and motivation was higher when female figures were used under women's role arousal. Lesser, Krawitz, and Packard (1963) also found that male figures aroused more n Achievement for all groups and for all conditions.

The white family structure in the United States is allegedly patriarchal (Deutsch, 1963), and the n Achievement findings in white female college students with respect to the sex characteristics of the figures used to assess such motivation is clearly consistent with identification of the male figure as the achiever in this society. The Negro sex role identification cannot be assumed to be identical to that of the average white person. Thus, the Negro female may not see the male figure as achievement related. Dating from slavery, society has reinforced with consistency and with strength the impoverishment of the Negro male's status. "The Negro boy does not generally have the opportunity to identify with a male figure who has a history of reinforcement for accomplishment (Deutsch, 1963, p. 66)." Thus, it is possible to hypothesize, on the basis of the cultural stereotype, that Negro female college students would demonstrate reverse effects from white female college students as a result of the manipulation of the sex of the figures used to study need for achievement. However, the prediction that Negro females would respond to female figures with greater n Achievement imagery than they would to male figures is not compelling or unambiguous. Based on recent societal events (civil right laws, riots, equalization of opportunities), it is entirely possible that the Negro middle-class female will see the male as the potential achiever in American society, at least in fantasy. The effects of the racial variable, Negro versus white figures, and the combination of the racial and sex variables could not be predicted with any surety because of conflicting evidence in the literature. The purpose of the study was to explore the effects of the cue character-

istics of race and sex separately, and in combination, on need for achievement scores in Negro college students.

METHOD

Subjects

All subjects were Negro college students from a Negro college in the South. All subjects would be classified as middle socioeconomic status because the orientation of the college, just as well as the middle class occupational orientation of the subjects. Most of the subjects were from Southern states, typical of the student body as a whole.

Procedure

Sixty-seven unpaid volunteers were randomly assigned to four conditions, each one differing in the stimulus characteristics of the figures presented: Negro male, Negro female, white male, and white female. The procedure was identical in all four conditions, the only variation being in the manipulation of the sex and the race of the figures used to elicit n Achievement imagery.

The achievement pictures were group administered by a Negro female experimenter under "relaxed" conditions. The pictures used were standard Murray TAT pictures suggested by Easter and Murstein (1964) as being pictures more likely to elicit n Achievement than other TAT pictures. The pictures were shown in the following order for all four groups: 1 (boy with violin), 2 (girl on farm holding books), 6B (old woman looking out of window with young man appearing sheepish), 8 BM (operation scene), 7 BM (father and son), 20 (ambiguous figure under street light). The change in racial and sexual characteristics of all figures but 20 was achieved by replacing the critical parts, for example, white face, with a new part. The changed features were glued on the original card using rubber cement.² The critical features could then be removed and replaced for use in the other experimental conditions. Thus, the picture remained the same with the exception of those features necessary to change the sex and race of the figure.

The instructions given to all four groups were identical. The particular instructions used were those instructions suggested by Atkinson (1958, p. 837). The pictures were presented for 20 seconds on a screen and then removed. Four minutes were allowed for writing each story.

The six stories were scored according to McClelland's scoring system "C" (Atkinson, 1958) in which it is possible to attain a maximum score of +11. For each individual n Achievement scores were summated across all six subscores. A productivity measure, that is, word count, was scored for each subject, counting the total number of words for all six pictures.

² The T-TAT was not used as it was judged particularly stereotyped, for example, New Orleans Creole angularity of features; the T-TAT did not vary sex of the figures; and the expressions plus situations were different from the Murray TAT.

TABLE 1

SCORES FOR RACIAL AND SEXUAL CHARACTERISTICS OF FIGURES: ANALYSIS OF VARIANCE OF NEED FOR ACHIEVEMENT SCORES

Source	Uncorrected			Corrected	
	df	MS	F	MS	F
Race (R)	1	294.12	8.08*	141.60	4.66
Sex (S)	1	765.47	11.02**	11.18	8.19**
R x S	1	67.75	1.17	91.15	1.09
Within	63	58.14		58.96	

* df for both corrected and uncorrected

* $p < .05$

** $p < .01$

Interscorer reliability was assessed by comparing independent scorings by the two authors on two cards. The average percentage of agreement for the two cards was 97%, and the average rank-order correlation coefficient for the two cards was .94.

RESULTS

Table 1 shows the effects of race and sex of the figures on corrected and uncorrected *n* Achievement scores.² The variance analysis of uncorrected *n* Achievement scores shows that there were two main effects, sex and race. Negro females scored higher on *n* Achievement measures for Negro rather than white figures and for male rather than female figures when the considered scores were uncorrected for length of protocol.

Since at least one of the conditions, race, affected productivity as well as *n* Achievement, the correlation between length of protocols and *n* Achievement scores for the total sample was obtained, as suggested by Atkinson (1958). A product-moment correlation of .42 was obtained, significant at the .001 level.

Because the significant correlation between protocol length and *n* Achievement scores indicates that the relationship between race of the figure and *n* Achievement scores may be spuriously affected by length of the protocol (Atkinson, 1958; Block, 1964; Veroff, Atkinson, Feld, & Gurin, 1960), the *n* Achievement scores were corrected to remove the influence of gross differences in length of protocol using the method suggested by Walker and Atkinson (1958). This particular method involved dividing the protocols into five subgroups according to length and then computing *t* scores ($M = 50$, $SD = 10$) for the need for achievement scores within each group of approximately equal length. The resulting analysis of variance on corrected *n* Achievement

² The distribution of scores eliminates the possibility that a few extreme scores are responsible for the results.

TABLE 2

ANALYSIS OF VARIANCE OF PRODUCTIVITY SCORES FOR RACIAL AND SEXUAL CHARACTERISTICS OF FIGURES

Source	df	MS	F
Sex (S)	1	114.00	12.11***
Race (R)	1	114.00	12.11***
S x R	1	114.00	12.11***
Within	63	114.00	

*** $p < .001$

ment scores, as shown in Table 1, indicates that the sex variable is unaffected by correcting the *n* Achievement scores. However, the effect of the racial variable is removed by the correction for protocol length.

Table 2 presents an analysis of variance of productivity scores instead of *n* Achievement scores, to test the effects of the various conditions sex and race, on verbal productivity. In contrast to Table 1 which represents *n* Achievement scores, Table 2 indicates that only the racial variable was effective, that is, the subjects were more productive verbally to pictures of Negroes than to pictures of whites irrespective of sex which is what Thompson found with southern Negroes.

The means for each of the four conditions when using uncorrected *n* Achievement scores, productivity scores, and corrected *n* Achievement scores are presented in Table 3.

DISCUSSION

The sex of the figure used in eliciting need for achievement themes does appear to be a factor for Negro females. The Negro female subjects were more responsive to pictures of males irrespective of race, in terms of producing achieve-

TABLE 3

MEAN SCORES ON DEPENDENT MEASURES AS A FUNCTION OF SEX AND RACE OF FIGURES IN SIX PICTURES

Condition	Uncorrected <i>n</i> Achievement	Productivity length	<i>n</i> Achievement scores corrected for protocol length ($M = 50$, $SD = 10$)
Negro male ($N = 16$)	9.80	556.93	56.40
White male ($N = 16$)	4.60	491.81	50.92
Negro female ($N = 16$)	2.00	542.69	47.18
White female ($N = 16$)	-7.36	451	47.10

ment-related themes. Although there was no interaction, the Negro male figure in uncorrected scores did elicit significantly more achievement themes than under the other three conditions. The same tendency to see the Negro male in terms of achievement themes was also present in the corrected scores, although in this case the racial variable did not significantly affect the results. Thus, one could conclude that, in spite of matriarchal orientation, the Negro male is seen by Negro college girls as a potential achiever, albeit in fantasy. This finding may be a function of the nature of the sample, essentially middle class, and it is intended to be followed up on lower socioeconomic class samples of Negro females to see if the expression of an Achievement to Negro male pictures will be generalizable. Possibly, the stereotype of the ineffectual Negro male is changing. The significance of the sexual variable is congruent with the work done by French and Lesser (1964), and others which shows that white females express more an Achievement to figures of males than to figures of females, unless specifically aroused in terms of woman's role.

One of the problems with respect to the racial variable is whether correcting an Achievement scores for productivity or length of protocol eliminates a significant aspect of what one is attempting to study. Block (1964) suggests that an Achievement scores contaminated by verbal productivity show more "sensible" relationships than either a corrected an Achievement score or a verbal productivity score. A tentative hypothesis is that Negro females are more stimulated to produce themes elicited by Negro pictures, but the nature of the theme is not necessarily achievement oriented. It is suggested that the Negro figures elicit more varied themes (e.g., dependency and love), whereas the white figures elicit a more constricted range of themes, which can account for greater productivity to Negro figures but not greater an Achievement. The findings that Negro figures do elicit more material (length of protocol) than do white figures is consistent with Thompson's original study and not consistent with those authors (e.g., Schwartz, Riess, & Cottingham, 1951) who found no effect of the race of the figures on productivity scores.

The results of this study indicate that the sex and race of the figures used to assess motivation in Negro female college students is important; Negro figures elicit more material than white figures, and male figures elicit more an Achievement than do female figures. Since Murray took sex into account and has separate sets of cards for men and women, the further elaboration of

race matched with the subject's race has value, as Thompson originally suggested, for clinical and research purposes, particularly with southern populations, more so, perhaps, than sex. More generally, this study implies that although the Negro male may be culturally devaluated, in fantasy he is seen as striving toward achievement-related goals.

REFERENCES

- ATKINSON, J. W. (Ed.) *Motives in fantasy, action, and society*. Princeton, N. J.: Van Nostrand, 1958.
- BLOCK, J. Some observations regarding the use of projectives in social and industrial research. *Journal of Social Psychology*, 1964, **64**, 157-160.
- COOK, R. A. Identification and ego defensiveness in thematic apperception. *Journal of Projective Techniques*, 1953, **17**, 312-319.
- DEUTSCH, M. M. Minority group and class status as related to social and personality factors in scholastic achievement. In M. M. Grossack (Ed.), *Mental health and segregation*. New York: Springer, 1963. Pp. 64-75.
- EASTER, L. V., & MURSTEIN, B. I. Achievement fantasy as a function of probability of success. *Journal of Consulting Psychology*, 1964, **28**, 154-159.
- FRENCH, E. G., & LESSER, G. S. Some characteristics of the achievement motive in women. *Journal of Abnormal and Social Psychology*, 1964, **68**, 119-128.
- KORCHIN, S. J., MITCHELL, H. E., & MELTZOFF, J. A critical evaluation of the Thompson thematic apperception test. *Journal of Projective Techniques*, 1950, **14**, 445-452.
- LESSER, G. S., KRAWITZ, R. N., & PACKARD, R. Experimental arousal of achievement motivation in adolescent girls. *Journal of Abnormal and Social Psychology*, 1963, **66**, 59-66.
- MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. I. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
- MERBAUM, A. D. Need for achievement in Negro and white children. *Dissertation Abstracts*, 1962, **23**, 693-694.
- MINGIONE, A. Need for achievement in Negro and white children. *Journal of Consulting Psychology*, 1965, **29**, 108-111.
- NUTTALL, R. L. Some correlates of high need for achievement among urban northern Negroes. *Journal of Abnormal and Social Psychology*, 1964, **68**, 593-608.
- RIESS, B. E., SCHWARTZ, E. K., & COTTINGHAM, A. An experimental critique of assumptions underlying the Negro version of the TAT. *Journal of Abnormal and Social Psychology*, 1950, **45**, 700-709.
- ROSEN, B. C. Race, ethnicity, and the achievement syndrome. *American Sociological Review*, 1959, **24**, 47-60.
- SCHWARTZ, E. K., RIESS, B. E., & COTTINGHAM, A. Further critical evaluation of the Negro version of the TAT. *Journal of Projective Techniques*, 1951, **15**, 394-400.

THOMPSON, C. E. The Thompson modification of the thematic apperception test. *Journal of Projective Techniques*, 1949, 13, 469-478.

VEROFF, J., ATKINSON, J. W., FELD, S. C., & GURIN, G. The use of thematic apperception to assess motivation in a nationwide interview study. *Psychological Monographs*, 1960, 74(12, Whole No. 499).

VEROFF, J., WILCOX, S., & ATKINSON, J. W. The achievement motive in high school and college age

women. *Journal of Abnormal and Social Psychology*, 1953, 48, 108-119.

WALKER, E. L., & ATKINSON, J. W. The expression of fear-related motivation in thematic apperception as a function of proximity to an atomic explosion. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton, N. J.: Van Nostrand, 1958. Pp. 143-159.

(Received March 16, 1966)

Journal of Personality and Social Psychology
1967, Vol. 5, No. 2, 249-252

DOGMATISM AND RESPONSES TO OPINION-CONSISTENT AND OPINION-INCONSISTENT INFORMATION

ROBERT E. KLECK¹ AND JERRY WHEATON

Williams College

The relationship between open- and closed-mindedness, as measured by the Dogmatism Scale, and responses to information consistent and inconsistent with own opinion was examined. As expected, dogmatic Ss did show less recall of inconsistent information and a greater tendency to evaluate consistent information more positively than did open Ss. The expectations that dogmatic Ss would show a greater preference than open Ss for information consistent with their opinions and would evaluate opinion-inconsistent information less favorably were not confirmed, though the data were in the anticipated direction. Consistent with some previous results in this area, Ss as a group did demonstrate marked preference for opinion-consistent information.

A large body of research data has accumulated suggesting that persons tend to arrange the elements of their cognitive systems in such a way as to minimize inconsistency. While considerable experimental work has focused on the general mechanisms by which cognitive inconsistency is avoided or reduced, little attention has been given to individual differences which may affect these processes. Both Festinger (1957) and Brehm and Cohen (1962) have emphasized the need for research on personality variables in this area.

A logical first step is to examine personality dimensions which are explicitly related to or descriptive of cognitive functioning. An obvious choice in this regard is the dogmatism dimension delineated by Rokeach (1960). According to his conceptualization, a person's belief-disbelief system, which includes all propositions one holds to be true or false, can be described as open to the extent that it is regulated by the need to know and understand, and closed to the extent that it is controlled by the need to defend against anxiety or threat. We might expect on the basis of this and other features of the dogmatism dimension that the more closed-minded person

would be both more disturbed by cognitive inconsistency and more motivated to avoid or minimize that inconsistency. In a recent study which is relevant to this proposition, Fillenbaum (1964) examined the relationship between dogmatism and dissonance reduction in a dissonance-arousal situation similar to that used by Aronson and Mills (1959). Fillenbaum expected that high-dogmatic subjects would be more attracted to a group when they had undergone an unpleasant initiation into the group than would low-dogmatic subjects. The expected positive relationship between closed-mindedness and dissonance reduction, as indexed by attraction to the group, was obtained. The results were rendered inconclusive, however, by the fact that in the control condition, in which subjects did not undergo a severe initiation, there was an even higher correlation between degree of dogmatism and valuation of the group.

The present design is concerned both with voluntary exposure to information which is perceived to be consistent or inconsistent with a person's attitudes and with the cognitive processes which are activated subsequent to involuntary exposure to consistent and inconsistent information. On the basis of Rokeach's conceptualization of open versus closed cognitive systems

¹ Now at Dartmouth College.

and previous research in this area, the following hypotheses were formulated:

Hypothesis A. Closed-minded subjects will show a greater preference for information consistent with their opinions than will open-minded subjects.

Hypothesis B. When exposed to opinion-consistent and opinion-inconsistent information, the closed-minded subjects will (a) recall less of the opinion-inconsistent information than the open-minded subjects, and (b) evaluate the opinion-inconsistent information less favorably and the opinion-consistent information more favorably than the open-minded subjects.

A variable which was anticipated to have effects similar to those of closed-mindedness as defined by Rokeach is strong commitment to, or involvement with, an opinion; that is, the more committed one is to a point of view, the more likely it may be that the toleration for inconsistency created by new information will be reduced. As part of the present design, therefore, we sought to explore the extent to which commitment reinforces the processes expected to characterize a closed cognitive system when it is exposed to consonant and dissonant information. The method adopted to manipulate increased involvement was that of having the subjects publicly commit themselves to an opinion position (Cervin, Joyner, Spence, & Heinzl, 1961; Rosenbaum & Zimmerman, 1959).

METHOD

The subjects were male* and female juniors recruited from a regional high school in western Massachusetts to participate in a study of "certain teen-age attitudes and opinions." The subjects completed the experiment in three separate sessions.

In Session I, the experimenter introduced the study to the students and administered Rokeach's Dogmatism Scale (Form E, consisting of 40 items). This was followed by a teen-age attitude questionnaire which asked for opinions on five issues relevant to the current teen-age culture. The issue chosen for experimental focus was that of the age at which drivers' licenses are issued. This issue was previously determined to be one of high salience and one on which we could expect high agreement. The other items included attitudes toward popular music, going steady, clothes, and delinquency.

In Session II, which took place 1 week later, one-half of the high-dogmatic subjects and one-half of the low-dogmatic subjects were randomly selected to undergo a public commitment procedure. Each subject appeared before a well-dressed college student who asked the subject to verbalize his feelings on the issues of popular music and the minimum driving age. Subsequently, all subjects were given the choice of reading additional information relevant to the driving-age issue which either favored the status

quo (16 years), or which favored raising the minimum age level (18 years). This information was contained in two fictitious newspaper articles carefully structured to be equal in strength of argument and in factual documentation. The opinion espoused in each of these articles was clearly indicated by their titles. The experimenter explained that there would probably not be enough copies of the articles to go around, and therefore the subjects should choose the one they most wanted to read. After the choice had been made, the experimenter reported that a recount of the copies of the two articles indicated that there would be sufficient numbers for everyone to read both. Copies were distributed so that one-half of the subjects received the preferred article and one-half received the nonpreferred article first. Both articles were followed by four evaluative scales and required each subject to express his opinion as to (a) how informed the author was, (b) how clear the arguments were, (c) how biased the author was in his approach, and (d) how valid the conclusions were. As soon as the subject had read and evaluated the first article he was given the second.

Two weeks later, in Session III, all subjects were given a test "to see what they remembered from the articles on teen-age driving." This test was composed of multiple-choice items based on the two articles with items randomized in regard to the article source. The general nature of the experiment was outlined to the subjects, and their reactions to the various procedures were solicited.

RESULTS AND DISCUSSION

The data analysis which follows is based on the 72 subjects who completed all three sessions of the experiment. The split on the Dogmatism Scale was made at the median resulting in a mean closed-minded score of 180.11 and a mean open-minded score of 131.31. The overall mean for the 72 subjects was 155.71, indicating that the sample was slightly more closed-minded than the groups on which Form E was originally developed (Rokeach, 1960).

It had been hypothesized that high-dogmatic subjects would show a greater preference for opinion-consistent information than would low-dogmatic subjects. These results are in the expected direction ($\chi^2 = 2.17$, $df = 1$), but they do not reach an acceptable level of significance. While differences in dogmatism do not show a significant relationship to information preference, the subjects as a group do show a marked preference for opinion-consistent information ($\chi^2/8 = 2.36$, $p < .02$). In a recent review of the literature, Freedman and Sears (1965) pointed out that the experimental results regarding selective exposure to opinion-consistent information are highly ambiguous. Several studies have shown a general preference for opinion-consistent information (e.g., Ehrlich, Guttman, Schönbach, &

TABLE 1
MEANS AND STANDARD DEVIATIONS
FOR RECALL SCORES

Dogmatism	Opinion consistent		Opinion inconsistent	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Open-minded	2.92	1.30	3.22	1.34
Closed-minded	3.47	1.26	2.53	1.34

Mills, 1957), several a preference for opinion-inconsistent information (e.g., Feather, 1962), and several no clear preference (e.g., Jecker, 1964). Freedman and Sears suggest that what has been demonstrated by these studies is that across situations difference factors are operative in determining information preferences. It is impossible to determine which factors conditioned a general preference for opinion-consistent information in the present situation. The above authors note that a particular type of information preference may depend on the specific attitudinal position. This theory is not true for the present data in that subjects on both sides of the issue demonstrated a preference for information consistent with their position on the issue. Finally, it should be noted that the public commitment variable did not show the relationship to information preference that had been anticipated.

Part 1 of Hypothesis B suggested that closed-minded subjects would demonstrate less recall of information not consistent with their opinions than would open-minded subjects. The relevant data are indicated in Table 1, and the analysis of variance in Table 2. The hypothesis is clearly supported in that the closed-minded subjects do recall less from the dissonant article than do the open-minded subjects ($t = 2.22$, $df = 70$, $p < .05$). This result supports Rokeach's notion that the closed-minded person is less able to integrate new beliefs into his cognitive system in

TABLE 2
ANALYSIS OF VARIANCE OF RECALL SCORES

Source	<i>MS</i>	<i>F</i>
Between Ss		
Open-closed (A)	0.17	
Public commitment (B)	2.51	1.39
A × B	0.84	
Error	1.81	
Within		
Consistent-inconsistent (C)	3.67	2.10
A × C	14.06	8.03*
B × C	1.17	
A × B × C	0.56	
Error	1.75	

* $p < .01$, $df = 1/68$.

TABLE 3
MEANS AND STANDARD DEVIATIONS FOR
ARTICLE EVALUATION SCORES

Dogmatism	Opinion consistent		Opinion inconsistent	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Open-minded	16.28	3.48	14.06	3.22
Closed-minded	18.47	2.98	13.22	3.38

that what is not recalled cannot be integrated. Rokeach (1960) suggested a similar recall deficiency for closed-minded subjects in regard to the new cognitive elements necessary to solve the Doodlebug problem. It is clear from his data and from a study conducted by Ehrlich (1955) that this decreased recall cannot be attributed to a difference in intelligence. Our own data support this in that if one compares recall scores for both opinion-consistent and opinion-inconsistent articles, the open- and closed-minded subjects do not differ in total items recalled. This is because the inferiority in recall experienced by closed-minded subjects for opinion-inconsistent items is balanced by a superiority for recall of opinion-consistent items, though this latter result does not reach an acceptable level of significance ($t = 1.80$, $df = 70$). In any case, it is unlikely that the decreased recall for inconsistent information is related to differences in general memory ability or to intellectual functioning per se.

It should be noted that given the type of recall data collected, it is not possible to say at what point the deficit for recall of inconsistent information has taken place; that is, it is possible that the decreased recall was a function of inattention to dissonant information while reading, rather than a memory loss experienced over time. This question needs to be explored further.

Part 2 of Hypothesis B proposes that closed-

TABLE 4
ANALYSIS OF VARIANCE OF EVALUATION SCORES

Source	<i>MS</i>	<i>F</i>
Between Ss		
Open-closed (A)	16.67	1.82
Public commitment (B)	11.67	1.28
A × B	15.34	1.68
Error	9.14	
Within		
Consistent-inconsistent (C)	502.51	39.48**
A × C	82.51	6.48*
B × C	5.84	
A × B × C	18.06	1.42
Error	12.73	a

* $p < .05$, $df = 1/68$.

** $p < .001$, $df = 1/68$.

closed subjects who evaluated opinion-consistent information more favorably and opinion-inconsistent information more unfavorably than did open-minded subjects. The data are summarized in Table 3, and the relevant analysis of variance is indicated in Table 4. Two effects reach an acceptable level of significance. In the first place, subjects as a group evaluate the opinion-consistent information more favorably than they do the opinion-inconsistent information ($F = 39.48$, $p < .01$). This result is similar to that found by several other researchers (e.g., Feather, 1963). It is possible that in the present case this outcome was a function of one article being better written than the other. That is, most subjects agreed with the opinion that the minimum driving age should not be raised and if the article taking this point of view was in fact a more convincing bit of writing it would account for the result. This alternative explanation is controverted by the fact that those subjects who took the opposite view, that is, that the driving age should be raised, evaluated the article favoring this position more favorably. The number of subjects in this category was not sufficient to render a statistical test meaningful, but the differences are of the same general numerical magnitude as those observed in the group overall.

In any case, the exact equivalence of articles is not essential to the central concern, which is a comparison between the responses of open- and closed-minded subjects. Again, the expected difference between these two groups is significant as indicated by the significant interaction ($F = 6.48$, $p < .05$). When this interaction is examined in terms of simple effects, it is apparent that its significance is generated primarily by differences in the evaluation of opinion-consistent information, that is, closed-minded subjects evaluated the consistent article more positively than did open-minded subjects ($t = 1.99$, $df = 70$, $p = .05$). The expected difference for the evaluation of the dissonant article was in the expected direction, but did not reach an acceptable level of significance ($t < 1.0$).

It should be noted once again that the expectation formulated regarding the effects of public commitment failed to materialize. In retrospect, the most conservative explanation may be that the public commitment manipulation was in fact a very weak one; that is, the audience consisted of only a single well-dressed college student to whom the subject verbalized his opinion. There is also a second factor which may be viewed as mitigating the effectiveness of the manipulation. It was anticipated that the effect of publicly committing oneself would serve essentially to in-

crease one's involvement with a given attitude position. We were however using an audience case driving age with which the subjects were already highly involved, and it is possible that they were involved to such a degree that a public statement of the opinion resulted in an insignificant increment in involvement.

The results, taken as a whole, do suggest that dogmatism as defined by Rokeach is systematically related to responses to opinion-consistent and opinion-inconsistent information. The effect comes not in terms of differences in selective exposure to that information, which was one of the experimental expectations, but rather in terms of the cognitive processes initiated subsequent to that exposure.

REFERENCES

- ARONSON, E., & MILLS, J. The effects of severity of initiation on liking for a group. *Journal of Abnormal and Social Psychology*, 1959, 59, 177-181.
- BREHM, J. W., & COHEN, A. R. *Explorations in cognitive dissonance*. New York: Wiley, 1962.
- CERVIN, V. B., JOYNER, R. C., SPENCE, J. M., & HEINZ, R. Relationship of persuasive interaction to change of opinion in dyadic groups when the original opinions of participants are expressed privately and publicly. *Journal of Abnormal and Social Psychology*, 1961, 62, 431-432.
- EHRLICH, H. J. Dogmatism and intellectual change. Unpublished master's thesis, Ohio State University, 1955.
- EHRLICH, D., GUTTMANN, I., SCHÖNBACH, P., & MILLS, J. Postdecision exposure to relevant information. *Journal of Abnormal and Social Psychology*, 1957, 54, 98-102.
- FEATHER, N. T. Cigarette smoking and lung cancer: A study of cognitive dissonance. *Australian Journal of Psychology*, 1962, 14, 55-64.
- FEATHER, N. T. Cognitive dissonance, sensitivity, and evaluation. *Journal of Abnormal and Social Psychology*, 1963, 66, 157-163.
- FESTINGER, L. *A theory of cognitive dissonance*. Stanford: Stanford University Press, 1957.
- FILLENBAUM, S. Dogmatism and individual differences in reduction of dissonance. *Psychological Reports*, 1964, 14, 47-50.
- FREEDMAN, J. L., & SEARS, D. O. Selective exposure. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 2. New York: Academic Press, 1965. Pp. 57-97.
- JECKER, J. D. Selective exposure to new information. In L. Festinger (Ed.), *Conflict, decision, and dissonance*. Stanford: Stanford University Press, 1964. Pp. 65-81.
- ROKEACH, M. *The open and closed mind*. New York: Basic Books, 1960.
- ROSENBAUM, M. E., & ZIMMERMAN, I. M. The effects of external commitment on response to an attempt to change opinions. *Public Opinion Quarterly*, 1959, 23(2), 247-254.

(Received March 23, 1966)

GALVANIC SKIN RESPONSES AND PREJUDICE¹

GARY W. PORIER AND ALBERT J. LOTT

University of Kentucky

In this experiment the consistency between the cognitive and emotional components of negative attitudes (prejudice) is investigated. A study by Rankin and Campbell which found greater GSR levels in white Ss to incidental hand contacts by a single Negro E than by a single white E was replicated employing a sample of Negro and white stimulus persons. The California F Scale and Rokeach's Opinionation Scale were used as self-report measures of prejudice. The Rankin and Campbell finding was not replicated. F Scale scores, but not Opinionation scores, were found to correlate significantly with GSR bias scores. Results are discussed in terms of ethnic prejudice, belief prejudice, and need for employing groups of individuals where person stimuli are used as independent variables.

While some support can be found in the literature for consistency among the components of attitudes (e.g., Campbell, 1947; deFleur & Westie, 1958; Harding, Kutner, Koslansky, & Chein, 1954), Secord and Backman (1964) state that

... sufficient empirical evidence is not yet available to support of consistency among affective, cognitive, and behavioral components of prejudice. Moreover, moves toward consistency may in many instances be counteracted by strong emotional anchorage of the affective component. Because prejudice involves strong emotions it should be tied to those physiological functions associated with the emotions (pp. 417-418).

Investigations of such physiological functions have tended to focus on galvanic skin responses (GSR) resulting from *indirect* exposure to the object of prejudice. For example, in a series of studies by Cooper and his associates (Cooper, 1959; Cooper & Pollock, 1959; Cooper & Siegel, 1956; Cooper & Singer, 1956), GSRs were recorded for subjects when a complimentary statement was read about a group against whom they were strongly prejudiced. GSRs to these statements were greater than to similar statements about groups toward whom subjects felt less antipathy. Similarly, when GSRs of white subjects viewing colored slides of Negroes in various social contexts were recorded in a study by Westie and deFleur (1959), prejudiced subjects were found to give larger GSRs than nonprejudiced subjects.

¹This paper is based upon the first author's thesis for the MA degree at the University of Kentucky.

While the above studies have provided physiological evidence for the emotional involvement in prejudice, they did not use a face-to-face situation between a prejudiced subject and the object of his prejudice. An investigation by Rankin and Campbell (1955) did just this and found a significantly greater level of GSR to a Negro than to a white experimenter on the part of white male subjects. In their experimental situation (ostensibly a study of word associations and anxiety) the Negro and white experimenters made incidental hand contacts with a subject while his GSR was recorded. Verbal attitudes toward Negroes were measured by paper-and-pencil tests in separate classroom sessions, but the correlation between attitude test scores and the GSR measure yielded inconclusive results. Since the Rankin and Campbell investigation represents a kind of ideal approach to studying the emotional component of prejudice, it was the intent of the present study to replicate it, while making use of an improved design. The methodological problems associated with the Rankin and Campbell study and the steps taken in the present study to reduce or eliminate them are listed below:

1. Rankin and Campbell used only one Negro and one white experimenter to provide the stimulus contacts for the subjects. The present replication utilized a sample of Negro and white experimental assistants. The need for sampling stimulus populations, especially characteristics of experimenter stimuli, was early recognized by Brunswik (1947), and

more recently by Hammond (1954) and McGuigan (1963). Regarding the Rankin and Campbell study, specifically, Masling (1960) has noted that "... since he [Negro experimenter] was 9 years older, 2½ inches taller and 27 pounds heavier than the white E, there was no conclusive proof that the difference, [in the subject's GSR] was a function of skin color [p. 75]."

2. In the Rankin and Campbell study the white experimenter had responsibility for all preliminary instructions and operations (attaching the electrodes), perhaps allowing time for adaptation to occur. The simple novelty or surprise of the appearance of another person, Negro or white, could, perhaps, have confounded their results. In the present experiment all preliminary manipulations were handled by an experimenter who took no part in the later experimental contacts.

3. Some subjects took their verbal attitude tests prior to the experiment and some afterwards in the Rankin and Campbell study. It was found that those subjects who took them after the GSR session were significantly more hostile than those who took them before. By having all subjects take the attitude test from 1 to several weeks prior to the experiment proper, this interaction was avoided in the present study, while at the same time separating the two situations by a relatively long time interval.

4. Of the direct and indirect verbal attitude tests used by Rankin and Campbell, only the former had any relationship to the GSR. Only direct attitude measures were utilized in this study.

5. The word lists used in the present study were composed of words with known emotionality based on previous GSR studies (Smith, 1922), whereas this information was not available for the words used by Rankin and Campbell.

With the above points in mind, an altered replication was carried out using the GSR as the measure of emotion, the California E Scale (Adorno, Frenkel-Brunswick, Levinson, & Sanford, 1950; Chapman & Campbell, 1959), and Rokeach's (1960) Opinionation Scale as the verbal measures of attitude, and two different word lists, one made up of

emotion-provoking words and the other of neutral words.

The present study also tested the following prediction: a positive relationship will be found between GSR differences to physical contact by Negro and white experimental assistants and scores on both the E Scale and Opinionation Scale, and this relationship will be greater for the E Scale than for the Opinionation Scale since the former measures attitudes toward Negroes (and other ethnic minorities or outgroups) directly while the latter purports to measure a more "general intolerance" based primarily on belief systems.

METHOD

Subjects

A sample of 60 white males, drawn from approximately 237 male members of the introductory psychology laboratory sections and three other lower division psychology classes at the University of Kentucky, served as subjects. All were volunteers fulfilling a course requirement for participation in some experiment during the semester.

Apparatus

The GSR apparatus consisted of a dermohmmeter and an attached calibrated strip chart recorder. (A detailed description of the dermohmmeter and chart recorder can be found in Porier, 1963.)

The wires from the dermohmmeter to the subject were attached to a set of C-shaped plastic clamps which assured firm contact and a constant pressure through the use of electrode jelly. A dummy set of wires was attached to the subject's left hand in every case. The wires led from the hands, under the door, and to their respective attachments (the dermohmmeter and the table leg for the dummy wire) in the control room. The only other object in the experimental room besides the subject, chair, desk, and wired electrodes leading to the control room was a tape recorder on which a recording of the stimulus words used in the word-association test was played. The tape recording of the two lists of words was made to insure standardization of inflection and emphasis.

Procedure

A week or more before the GSR session, the E Scale and Opinionation Scale, in a combined form of 58 items, was administered to all the classes from which the subjects were later obtained.

The 18-item E Scale had half the items stated in a positive direction and the other half in a negative direction to control for an acquiescence response set (Chapman & Campbell, 1959). These items were alternated with 40 positively and nega-

tively stated items of the Opinionation Scale, American Version—Form C (Rokeach, 1960). To avoid any possible biasing effects in the GSR situation, the subjects' attitude scales were not scored until the experiment was completed.

In the experimental situation, each subject was met individually and asked to participate in a word-association test in which his emotional reactions to a list of words would be measured. The experimenter responsible for monitoring the equipment escorted the subject from a waiting room, through the control room, into the experimental room. After the electrodes were attached to subject's hands, the experimenter informed him that there would be a 10-minute waiting period and left the experimental room.

The GSR apparatus was turned on at this point to establish the subject's basal level of resistance. After an initial period of fluctuation, the resistance level generally became a slowly but steadily increasing positive function on the recording chart. When sufficient time for this level to be reached had elapsed (approximately 10 minutes), the same experimenter again entered the experimental room and read the following instructions:

This is an experiment to test your reactions to certain words. During the experiment these electrodes will remain attached to your hands. Please try to remain as still as possible, and especially try not to move your hands. Wait for one of the assistants to adjust and remove the apparatus. When the experiment begins, I shall turn on this recorded list of words and leave the room. Following each word you are to call out the first word you associate with the one you hear from the recorder. Are there any questions?

Please remain as still as possible until the experiment begins.

The experimenter turned on the tape recording of either the emotional or neutral word list, according to the condition being run at the time, and left the room. A clock-timer switch was thrown simultaneously to enable the experimenter to time the entrance of an experimental assistant so as to coincide with a blank space on the tape. The chart recorder switch was also turned on at this point to begin taking the record of the subject's GSR. The tape recordings of the word lists were divided into the following segments:

1. A 3-minute blank space to further allow the subject's base line of response to become well established after the experimenter's exit from the experimental room.

2. *First experimental contact.* A 1-minute blank period in which a white or Negro assistant entered the experimental room from the control room, made an "adjustment" on the dummy electrodes on the left hand (which was in closest proximity to the door from the control room), and left, being in the room approximately 20 seconds, as instructed.

3. *Second experimental contact.* Another 1-minute blank period in which a Negro or white assistant

went through the same procedure as the previous assistant.

4. The first 10 words from the word list, spaced at 20-second intervals to allow for the latency and return to prior level of the GSR.

5. *Third experimental contact.* Another 1-minute period of blank tape, during which the Negro or white assistant entered again and made another "adjustment" for 20 seconds.

6. The completion of the word list, during which the final four words were spaced at 20-second intervals as above.

7. *Fourth experimental contact.* The fourth and final experimental contact made by the white or Negro assistant following the same procedure as above.

8. A final 1-minute period of blank tape to allow the GSR level to return to the level prior to stimulation.

The recording chart was marked at the time an assistant entered the room and at the time he returned. For purposes of this experiment, and in line with the method employed by Rankin and Campbell, the dependent measure was the difference in GSR level between the time when an assistant entered the experimental room and the lowest point indicating the greatest reaction achieved before he left the room.

At the end of the experimental session the true purpose of the experiment was explained to each subject. Subsequent questions of a general nature about the experiment were answered, the importance of secrecy concerning the purpose of the experiment was stressed, and the subject was dismissed.

An attempt to control for age and size differences in the assistants was made by matching the Negro and white assistants for a particular experimental session as nearly as possible on these dimensions. A further attempt to minimize any differences in clothing and incidental appearance was made by having both assistants wear knee-length white lab coats during the experiment. Twelve Negro undergraduate, graduate, and medical students, and 21 white graduate students served as experimental assistants.

The experiment was run in two counterbalanced sequences. In Sequence A, the white assistant made the first and fourth contacts, and the Negro assistant the second and third (W-N-N-W). In Sequence B, the Negro assistant made the first and last contact, and the white assistant the second and third (N-W-W-N). The two sequences and the two word lists provide four possible combinations of conditions, as follows, with 15 subjects in each: (a) A-E (Sequence W-N-N-W, emotional word list),² (b) A-N (Sequence W-N-N-W, neutral word list),³ (c) B-E (Sequence N-W-W-N), emotional word

² Some examples of emotional words are love, flunk, divorce, and mother.

³ Some examples of neutral words are pencil, give, table, and pond.

list), and (d) B-N (Sequence N-W-W-N), neutral word list.

RESULTS

The suggestions made by Haggard (1947) for statistical treatment of GSR data were followed in this study. Each of the four dependent measure scores (as previously described) obtained on each subject was converted into a log change in conductance. This conversion of the GSR data is also similar to the procedure used by Rankin and Campbell (1955).

In order to reduce the GSR data to a single score for each subject (i.e., to obtain a measure of the difference in reaction to the Negro and white assistants), the sum of the two converted GSR scores to the white assistant was subtracted from the sum of the two converted GSR scores to the Negro assistant. This procedure produced a single GSR *bias* score for each subject.

Whether contacts by the Negro and white assistants were different with the emotional and neutral word lists was first tested by *t* test. The 30 GSR *bias* scores obtained under the emotional word list (Conditions A-E and B-E) were found not to differ significantly ($t = .07$) from those obtained under the neutral word list (Conditions A-N and B-N). Since varying the emotionality and neutrality of the word lists had no effect on the results, the data from these lists were combined in subsequent analyses.

Product moment correlations between the GSR *bias* scores and scores on the two verbal measures of prejudice were obtained. The predicted relationship between GSR and E. Scale scores was confirmed at the .01 level of significance ($r = .38$, $N = 60$), while the relationship between GSR and the Opinionation Scale failed to reach significance ($r = .13$, $N = 60$).

Present Results Compared with Those of Rankin and Campbell

A first attempt to determine whether a differential GSR reaction was made to the Negro and white assistants was carried out by a *t* test between the 60 combined GSRs to the two white contacts and 60 combined

GSRs to the two Negro contacts. The analysis indicated no significant difference between the two ($t = 1.31$, $p > .05$), and thus represents a failure to replicate the Rankin and Campbell (1955) finding.

The single GSR *bias* score of each subject in the present study is shown in Figure 1. These scores ranged from -1.60 to 1.55. Keeping in mind that a positive score means a greater overall reaction to the Negro as opposed to the white assistant, and noting the similarity to a normal distribution, the breakdown of positive and negative scores (32 plus, 28 minus) leads one to reject the possibility of a general differential reaction to Negroes on the part of the white subjects in this experiment. In the Rankin and Campbell study, however, 36 of 40 subjects had positive GSR *bias* scores; such a split has a probability of less than .0000001, while a 50-50 split represents chance expectancy.

To further compare the results of the present study with those of Rankin and Campbell, the mean GSR scores made to white assistants and those made to Negro assistants on each of the four contact occasions were plotted by sequence of contact (Sequence A, W-N-N-W; Sequence B, N-W-W-N) for both studies. This comparison is shown in Figure 2. (Some of the raw data for this figure came from Rankin, 1951.)

Rankin and Campbell found that the sequence of contact had a significant effect at less than the .01 level of significance. However, most of this effect was present in the initial contact with some adaptation present for Contacts 2 and 3 and the effect again being present on Contact 4. A *t* test for Contact 1 between the means of Sequence A

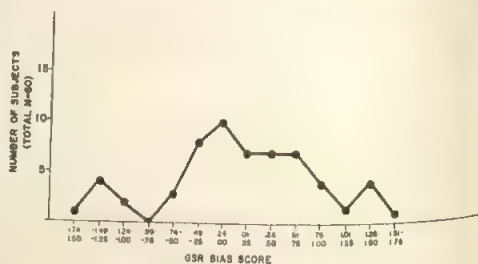


FIG. 1. Distribution of GSR bias scores (total sample).

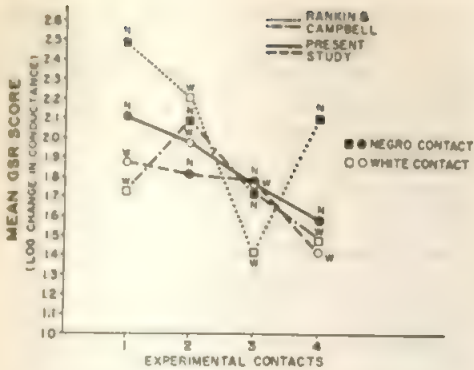


FIG. 2. Comparison of Rankin and Campbell's mean GSR scores (plotted by sequence) with the mean GSR scores of the present study.

(W-N-N-W) and the means of Sequence B (N-W-W-N) of the present study was non-significant ($t = .475$), and the remaining three between-condition mean comparisons show the means to be even closer together than those for Contact 1.

Comparing the curves representing the two sequences which were obtained from both studies, it can be seen that a comparatively steadily decreasing curve was found for both sequences in the present study, suggesting that no matter what sequence was used, a more definite continuous adaptation to the stimuli

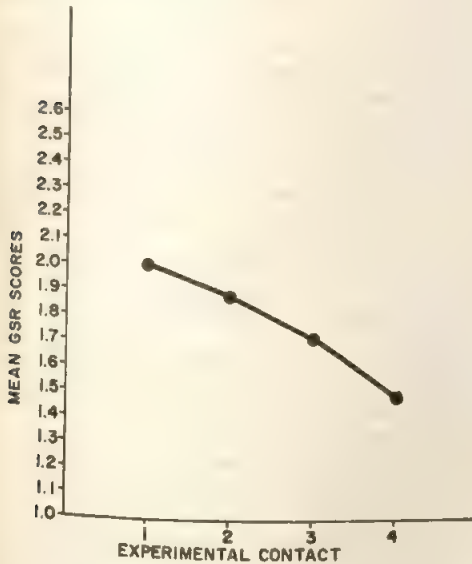


FIG. 3. Combined trial means illustrating the adaptation effect in the present study.

(Negro or white assistants) was taking place than that suggested by the Rankin (1951) and Rankin and Campbell (1955) data. The steadily decreasing function which results when the conditions are combined further suggests this adaptation effect. This function, using the total mean of each contact, is shown in Figure 3.

DISCUSSION

The low but significant correlation between the GSR *bias* scores and E Scale scores indicates that differential prejudice toward Negroes (and other minority groups), as inferred from this particular paper-and-pencil test, is associated with differential emotional reactions to Negroes and whites in a face-to-face behavioral situation. These results provide evidence for consistency between the cognitive and emotional components of negative attitudes. In the present study, affective and cognitive components of prejudice were measured and related by providing the subjects with actual contacts (a physical touch of the hand) by a sample of Negro and white experimental assistants. What the subjects said they would do or feel when responding in a paper-and-pencil situation was compared with their involuntary emotional reactions in an interpersonal situation. The correlation between E Scale and GSR is in agreement with data from Westie and De Fleur (1959), in which prejudiced subjects tended to give greater levels of autonomic response to color slides of Negroes than did an unprejudiced group.

The nonsignificant relationship obtained between responses to the Opinionation Scale and GSR *bias* scores warrants some comment.⁴ Rokeach (1960) has suggested that the Opinionation Scale is a measure of general prejudice under which racial prejudice, as a special case, may be subsumed. Under the conditions of the present investigation, the subjects had no differential information concerning the Negro and white assistants except that they differed in skin color. According to

⁴ In addition to the total opinionation score, right and left opinionation were correlated separately with GSR *bias* scores. Neither was significantly related to the GSR score (right $r = -.033$, left $r = .198$).

recent findings of Stein, Hardyck, and Smith (1965),

When subjects are forced to evaluate stimulus individuals in terms of their beliefs, then belief-congruence is more important than race. But when the belief component is not provided, spelled out in considerable detail, subjects will react in racial terms on the basis of assumptions concerning the belief systems of others, and of emotional or institutional factors [p. 289].

Since a nonsignificant correlation between Opinionation Scale scores and GSRs emerged in this study, where belief similarity or dissimilarity was not made experimentally salient, it appears that the authors' subjects did not assume major belief differences between the Negro and white assistants, and we are, therefore led to the conclusion that the subject sample was relatively low in racial prejudice. Support for this interpretation comes from three sources: (a) No significant differential GSR reaction was found to the Negro assistants as compared with the white assistants; (b) as a group, the subjects had relatively low mean scores on the E Scale;⁵ (c) the results of a study by Byrne and Wong (1962) showed that "highly prejudiced subjects were found to assume greater attitude dissimilarity between themselves and a Negro stranger than between themselves and a white stranger; for subjects low in prejudice the assumed dissimilarity scores for whites and Negroes did not differ [p. 253]." If the authors' subjects are regarded as low in prejudice, there is now no reason to expect them to have assumed greater belief dissimilarity between themselves and the Negroes than between themselves and the whites.

The discrepancy of the present results with those of Rankin and Campbell, in that no differential GSR was found to the Negro assistants as compared with the white assistants when all subjects were combined,

⁵ For the 60 experimental subjects the following scores were obtained: mean E Scale score = 63.22 out of a possible 126.00; mean score of anti-Negro items = 18.47 out of a possible 42.00, using a 7-point scoring system ranging from a low item score of 1 to a high item score of 7. The total pool of 237 means from which the subjects were drawn showed the following scores: mean E Scale score = 66.94 out of a possible 126.00; mean anti-Negro score = 20.09 out of a possible 42.00.

may best be explained, in the view of the present authors, by procedural differences in present person stimuli (Negro and white assistants) to the subjects (as well as the other controls mentioned earlier). Rankin and Campbell used only one Negro and one white stimulus person, whereas the present study utilized a group of each. When many individual difference characteristics are canceled out through a grouping procedure, it appears that the color of the assistant's skin was not significant enough to produce differential galvanic skin reactions in the type of subjects used in this experiment.

REFERENCES

- ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. K., & SANFORD, R. M. *The authoritarian personality*. New York: Harper, 1950.
- BRUNSWIK, E. *Systematic and representative design of psychological experiments*. Berkeley: University of California Press, 1947.
- BYRNE, D., & WONG, T. J. Racial prejudice, interpersonal attraction, and assumed dissimilarity of attitudes. *Journal of Abnormal and Social Psychology*, 1962, 65, 246-253.
- CAMPBELL, D. T. The generality of a social attitude. Unpublished doctoral dissertation, University of California, Berkeley, 1947.
- CHAPMAN, L. J., & CAMPBELL, D. T. The effect of acquiescence response-set upon relationships among the F scale, ethnocentrism, and intelligence. *Sociometry*, 1959, 22, 153-161.
- COOPER, J. B. Emotion in prejudice. *Science*, 1959, 130, 314-318.
- COOPER, J. B., & POLLOCK, D. A. Identification of prejudicial attitudes by the GSR. *Journal of Social Psychology*, 1959, 50, 241-245.
- COOPER, J. B., & SIEGEL, H. E. The GSR as a measure of emotion in prejudice. *Journal of Psychology*, 1956, 42, 149-155.
- COOPER, J. B., & SINGER, D. N. The role of emotion in prejudice. *Journal of Social Psychology*, 1956, 44, 241-247.
- DE FLEUR, M. L., & WESTIE, F. R. Verbal attitudes and overt acts: An experiment on the salience of attitudes. *American Sociological Review*, 1958, 23, 667-673.
- HAGGARD, E. A. On the application of analysis of variance to GSR data: I. The selection of an appropriate measure. *Journal of Experimental Psychology*, 1947, 51, 180-189.
- HAMMOND, K. R. Representative versus systematic design in clinical psychology. *Psychological Bulletin*, 1954, 51, 180-189.
- HARDING, J., KUTNER, B., PROSHANSKY, H., & CHEN, I. Prejudice and ethnic relations. In G. Lindzey (Ed.), *Handbook of social psychology*. Reading, Mass.: Addison-Wesley, 1954. Pp. 1021-1061.

- MASLING, J. The influence of situational and interpersonal variables in projective testing. *Psychological Bulletin*, 1960, **57**, 65-85.
- MCGUGAN, F. J. The experimenter: A neglected stimulus object. *Psychological Bulletin*, 1963, **60**, 421-428.
- PORIER, G. W. The relationship between the galvanic skin response and prejudicial attitudes. Unpublished master's thesis, University of Kentucky, 1963.
- RANKIN, R. E. The galvanic skin response as a physiological measure of social attitudes. Unpublished master's thesis, Ohio State University, 1951.
- RANKIN, R. E., & CAMPBELL, D. T. Galvanic skin-responses to Negro and white experimenters. *Journal of Abnormal and Social Psychology*, 1955, **51**, 30-33.
- ROKEACH, M. *The open and closed mind*. New York: Basic Books, 1960.
- SEBORD, P. F., & BACKMAN, C. W. *Social psychology*. New York: McGraw-Hill, 1964.
- SMITH, W. *The measurement of emotion*. New York: Harcourt, Brace, 1922.
- STEIN, D. D., HARDYCK, J. A., & SMITH, M. B. Race and belief: An open and shut case. *Journal of Personality and Social Psychology*, 1965, **1**, 281-289.
- WESTIE, F. R., & DE FLEUR, M. L. Autonomic responses and their relationship to race attitudes. *Journal of Abnormal and Social Psychology*, 1959, **58**, 340-347.

(Received May 26, 1966)

THE ACHIEVEMENT MOTIVE IN WOMEN: IMPLICATIONS FOR CAREER DEVELOPMENT¹

RHODA BARUCH

Center for Research in Careers, Harvard University

This study was suggested by McClelland's proposition that *n Achievement* is an important psychological variable behind social and economic change. 2 hypotheses concerning the changing patterns of employment of women were tested: (a) There is a temporal cycle in *n Achievement* associated with age and family situation; and (b) high *n Achievement* is associated with return to paid employment. Data from TAT stories written by 137 Radcliffe alumnae supported the hypotheses. A broader test with a nationwide sample of 763 women failed to confirm either hypothesis. The first relationship obtained for the college women in the sample. Further analysis indicated a time lag between increased *n Achievement* and increased participation in paid employment.

In the past 2 decades in the United States extraordinary changes in the employment patterns of women have occurred. The representation of women in the labor force, and particularly the proportion of married women, older women, and mothers, is greater than ever before in our nation's history. These changes are considered "the most important employment trend in the country today" by Assistant Secretary of Labor Esther Peterson (Fleming, 1964, p. 2) and have been termed the "silent revolution" in the role of women by the National Manpower Council (1958). The magnitude and something of the character of these changes may be conveyed by these statistics:

At present, one worker in three is a woman, whereas in 1940 approximately one worker in four was a woman, and in 1920 only one out of every five workers was a woman (President's Commission on the Status of Women, 1963, p. 28). As many as 30,000,000 women worked for some time during 1963 (Fleming, 1964); of these women, 24% were single, 55% were married, and 21% were widowed, divorced, or separated (Rosenfeld & Perrella, 1965, p. 1). Again, comparing the marital status of women in the labor force in 1940

with 1960, we find that at the earlier date 17% of the married women were working and only 31% of the working women were married, whereas 20 years later 32% of all married women were employed outside the home, and they comprised a majority (54%) of the population of working women (Degler, 1964).

These increases in the representation of married women in paid employment cannot be attributed only to the presence of young married women, although with more women marrying than ever before (93%) at an earlier age—20 is the median (Ginzberg, 1958, p. 313)—these factors do contribute to the rise of the proportion of married women in paid employment. Interestingly, the increasing participation of married women in the labor force has been accompanied by a marked increase in the participation of older women since 1940 (President's Commission on the Status of Women, 1963, p. 11). Currently the average age of the working women is 41 (Fleming, 1964), and the year 1963 found more than half of all women in the 45-54 age bracket in paid employment.

With increased participation of married, older women has come an increase in the number of mothers in the labor force to an estimated 10,000,000 (Nye & Hoffman, 1963). As one might expect, the presence of young children inhibits the participation of many women in paid employment, but the percentage of married women in the labor

¹ This article is based on a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Education at the Harvard Graduate School of Education, 1966. The writer wishes to express her appreciation to Anne Roe, who as advisor provided important assistance throughout the course of this project.

age rises with the age of the children (President's Commission on the Status of Women, 1963, p. 20).

Curiously, many social scientists have not only failed to predict these trends of continued increase in the employment of married women, but also have regarded the employment of mothers during World War II as "a temporary adaptation to a unique set of wartime conditions [Nye & Hoffman, 1963, p. 6]." Women, after the emergency was over, were expected (predicting from the prevalent psychological theories concerning women) to leave their jobs, renounce more active goals and roles, return to their homes, and fulfill their true, eternally feminine roles. As recently as 1944, Deutsch (1944, p. 386) wrote:

In this country during the present war incomparably wider strata of women are active in occupational fields . . . But the majority of women whom war has made more active than ever, will return as quickly and energetically as possible to the basically conservative because always dominant feminine experience, regardless of social and cultural upheavals.

The theory of achievement motivation of McClelland (1961) may help in understanding the changing patterns of the employment of women. McClelland's hypothesis that this particular motive is a psychological variable behind social and economic change has been lent support by a broad range of studies. There is evidence that the level of the achievement motive (*n* Achievement) of a society is significantly related to the presence or absence of full-time entrepreneurial activity in the culture (Child, Storm, & Veroff, 1958).

THE HYPOTHESES: TEMPORAL CYCLE IN *N* ACHIEVEMENT AND RETURN TO PAID EMPLOYMENT

Hypothesis I

The *n* Achievement of women is associated with their age and family situation. A woman's adult life may be considered to comprise three phases: one before the advent of children, one when home and children are her major concern, followed by one when the family has been established. Specifically, *n* Achievement, tested under neutral conditions, declines from the level of the first phase dur-

ing the second phase but during the third phase returns to earlier levels.

This hypothesis is suggested in part by Matthews' (1900, p. 141) proposition (2) which states that succeeding ascendancies of needs will mold women's life developmental pattern. Degler (1964) also writes of the "marked temporal cycle" with regard to women and careers. One expects the achievement motive in women to follow such a temporal cycle, with a decline during the period when women are most involved in the mother role and a subsequent return in the strength of this motive as their children leave the infancy stage. Parsons (Parsons & Bales, 1955) describes the differentiation of roles within the family, suggesting that the wife-mother has an expressive-integrative role to fulfill. A woman is not likely to construe her role vis-à-vis her infant in terms of being a "good mother" and competing with a standard of excellence in doing a "good job" in achievement terms, but rather in terms of affectionate interpersonal relationships and concern for the dependent infant. Nurturance and affiliation needs can be expected to predominate during this period. Erikson designates the period of adulthood as the period of generativity in the healthy personality. For a brief period, at least, we assume this generativity is very literally expressed in creating and caring for the next generation. However, for those with a strong long-term disposition toward achievement motivation we expect, after a moratorium on achievement striving during which the family has been created, a return of this motive to regnancy.

A second hypothesis derives in part from Roe (1962) who, in reflecting upon what an occupation outside the home may add to the total satisfaction of a woman's life, stresses the need for a feeling of competence and the need to receive open recognition and recognition from peers.

Hypothesis II

For those women whose families are established, high *n* Achievement will be associated with paid employment and return to work.

This hypothesis tests the association of *n* Achievement with work status or career pattern and attempts to explain the seeking out

of paid employment by the older women in our labor force in terms of achievement needs. There are parallels between the recent trend of women leaving the home for paid employment and the earlier changes in the patterns of men's work. Those changes occurred during major economic growth periods and were, according to McClelland's (1961) hypothesis, a result of increase in *n Achievement*.²

METHOD

Sample

Several considerations determined the choice of the sample for the first tests of the hypotheses and require a brief review of the difficulties encountered with *n Achievement* in women. The procedures for testing *n Achievement* had been developed with male college students as subjects. Using the same procedures which produce an increase or arousal in *n Achievement* for males, Veroff, Wilcox, and Atkinson (1953) could not produce a similar increase for high school or college female subjects. It would seem that the achievement motive score had no validity for the women subjects. Yet the performance by these female subjects on an anagrams task, similar to the scrambled words task used by Lowell (McClelland, 1961) as a validation technique, was found to correlate with *n Achievement* scores for the women, and these results were essentially the same as those obtained by Clark and McClelland with men. The superiority of the performance of the high *n Achievement* women was pronounced on this anagrams task. To add complexity to this puzzle, pictures containing male and female central figures elicited different responses in terms of level of *n Achievement*, with the male picture cues producing more achievement imagery in both sexes.

Angelini, in Brazil, attempting to replicate the experiments reported by Veroff et al. (1953), reported a significant increase in *n Achievement* scores of both men and women following experimental arousal conditions. These women were presumed to be highly competitive and placed great stress on intellectual accomplishment (Atkinson, 1958, p. 77). For women such as these Brazilian college women, the *n Achievement* scores could be used with the same amount of confidence as with male subjects. Lesser, Krawitz, and Packard (1963) argued that the failure to demonstrate increases in achievement motive in American women in response to experi-

mental conditions stressing intelligence and achievement is related to the fact that American subjects had been insufficiently concerned with standards of intellectual excellence and with achievement through the development of intellectual skills. Lesser et al. studied the highly selected, highly motivated Hunter College High School population and divided their sample into two groups, *achievers* and *underachievers*. Although the *achievers* were no more responsive to arousal conditions than *underachievers*, a highly significant second order effect was observed; the achievement-motive scores of the achieving girls increased in response to the usual achievement-oriented conditions when they were telling stories with pictures of females, but not when the pictures involved male central figures. In contrast, the achievement-motive scores of *underachievers* increased significantly in response to achievement-oriented conditions when they were producing stories with male central figures but not in response to pictures of females.

Another study by French and Lesser (1964) found a significant interaction between the value-orientation and arousal conditions. Those women subjects who placed high value on intellectual attainment scored higher on achievement motive only under the arousal conditions emphasized intellectual success. A further finding supported the hypothesis that a relationship exists between motivation and performance (in a validation task) when the cues correspond to the subject's valued goals, but not otherwise.

The first test of our hypotheses is made on a highly selected sample, the Radcliffe alumnae, who should constitute a group analogous to Lesser et al.'s (1963) "achievers" and Angelini's Brazilian college women. Again, these women should resemble those college women studied by French and Lesser (1964) who value intellectual attainment and leadership. The Radcliffe alumnae are not only highly selected but are also self-selected and have, by virtue of their gaining admission to the college and completing requirements for the bachelor's degree, demonstrated on the behavioral level concern with competition with a standard of excellence. In contrast with other able women who do not undertake such collegiate challenge, achievement striving was accepted by them as compatible with their feminine identity or as ego syntonic. Hence, the achievement motive measure for women we assume is here valid and appropriate.

Utilizing data obtained from a nationwide study of motivation, a second and broader test of the hypotheses was made subsequent to the Radcliffe study. A nationally representative sample of 768 women who had participated in a survey conducted by the Survey Research Center (Gurin, Veroff, & Feld, 1960; Veroff, Atkinson, & Feld, 1960) comprised the second sample.

Procedure

A series of six pictures was selected for presentation to the subjects. Three of these pictures were

²The United States has undergone, during the decades in which women's employment trends changed, a period of economic growth such that electrical energy output, a metric used by McClelland as an index of economic growth, increased almost sixfold in the period from 1940-1960 (Golen-paul, 1964).

from the female form (a series of six) and three were from the male form (again a series of six) which had been used in the assessment of motivation in the nationwide study² (Veroff et al., 1960).

² The pictures used in the Radcliffe study are described below, and both the figures for the percentage of stories coded for achievement imagery in the nationwide survey and in the Radcliffe sample are given. The pictures are purported to be a sampling from "universally relevant situations" to help minimize the bias in the measuring instrument.

1. Two women standing by a table, one of whom is working with test tubes: 45% Achievement imagery in nationwide sample; 83.5% Achievement imagery in Radcliffe alumnae sample.

2. Group of four women. One standing, the others seated facing each other: 10% Achievement imagery in nationwide survey; 15% Achievement imagery in Radcliffe sample. (This picture had high cue value for the affiliation motive, and 44% of the women in the nationwide survey told stories with affiliation as the theme.)

3. Woman kneeling and applying a cover to a chair: 36% Achievement imagery in nationwide survey; 80% Achievement imagery in Radcliffe alumnae sample.

4. Two men (inventors) in a shop working at a machine (2): 32% Achievement imagery in nationwide survey (male subjects); 76% Achievement imagery in Radcliffe alumnae sample.

5. Four men seated at table with coffee cups. One man is writing on sheet of papers (101): 23% Achievement imagery in nationwide survey (male subjects); 64% Achievement imagery in Radcliffe alumnae sample.

6. Man seated at drafting board (28): 13% Achievement imagery in nationwide survey (males); 69% Achievement imagery in Radcliffe sample.

The numbers in parentheses after each description refer to the list number in Atkinson (1958, pp. 832-841).

Although the comparison which is invited by juxtaposing the two figures for each picture is not completely justified because of differences in the methods of obtaining the stories, one can nevertheless see that the pictures have a high cue value for the Radcliffe sample. This sample is probably far more homogeneous and high in Achievement than the population in general. In all, of the 300 people approached, 178 responded with stories, yielding a return rate of 59%. A printer's error affected some of the material sent out in the second mailing. Inadvertently, some of the test booklets had only two questions per picture, each question repeated twice, instead of the usual prescribed four questions per picture for the TAT. No reasonable adjustment of scores for these pictures could be made for these data obtained with incorrect procedures. It was decided to delete those 41 subjects from the sample, and the second mailing thus resulted in the addition

Several of these pictures had also been used in Lewis' (1958) High School Study. In the nationwide survey, the pictures were, however, the women had been given only the female form. The nationwide test booklet also allowed somewhat less time than was given in the Radcliffe study.

The pictures were presented in random order so that the effect of position could be considered minimal. There are no data in this study to suggest that any consistent effects were the result of order of presentation. Veroff et al. (1960) used a fixed order of presentation, in that college picture first, then the series of picture presentation had no systematic effect on the total scores on stories.

From the Alumnae Directory of Radcliffe College (1962), a list was compiled of all graduates from each of the classes of '33, '34, '35, '36, and '37 living in the Boston area. To these 417 subjects, chosen at random from the geographic lists in the Alumnae Directory, a letter was sent. The letter was enclosed in a small envelope which in turn was affixed to a large manila envelope containing the test booklet.

The instructions printed in the large manila envelope were enclosed in the larger envelope. The test booklet was sealed with a small white sticker, and the larger envelope had a seal on it with an inscription requesting that the letter be read before the large envelope was opened. A "neutral" set was intended in the tone of the letter rather than any arousal conditions. Precautions such as the use of seals and the recording of time were taken in order to encourage spontaneity when the task was actually begun. A follow-up phone call to all subjects who had not responded within the first 7 weeks elicited many unscripted assurances from the subjects that, since they had not had the required time, they had not as yet even opened or peeked at the materials. The subjects seemed to understand and cooperate with these procedures, and, although some subjects apologized for taking more than the specified 30 minutes, few exceeded 45 minutes, and this period included time for the last questionnaire or calendar page.

Several subjects were temporarily out of the country, seriously ill, or moved without leaving a forwarding address. A small number sent the materials back unanswered, refusing without explanation or on the grounds that they were too busy or too sophisticated about the projective techniques used in the study. A larger number not only responded, but enclosed notes which expressed enthusiasm for only 58 scorable booklets, although the return rate was higher than for the first mailing.

An earlier analysis for a partial sample had shown that this association was not significant when one included the unmarried women. We wondered whether those women with low need achievement and a strong career pattern were working for "other reasons," that is, to support themselves because they were unmarried. Restricting our sample then to the married women, we found a significant relationship with career pattern for the married women. Here we restrict this analysis to the married women.

TABLE 1

ACHIEVEMENT MOTIVE IN YEARS AFTER COLLEGE OF RADCLIFFE ALUMNAE

n Achievement total score	Yrs. after graduation					
	5	10	15	20	25	Totals
High						
Observed	16	2	10	11	10	49
Expected	(10.014)	(8.226)	(10.729)	(9.299)	(10.729)	
Contribution to chi-square	.577	.4712	.049	.311	.049	
Medium and low						
Observed	12	21	20	15	20	88
Expected	(17.985)	(14.773)	(19.270)	(16.700)	(19.274)	
Contribution to chi-square	1.991	2.624	.027	.173	.027	
N	28	23	30	26	30	137

Note.— $\chi^2 = 13.544$, $df = 4$, $p < .01$, $N = 137$.

couragement to the investigator or interest in receiving the results of the study. The nature of any volunteer error operating here is not understood.

A total of 82 scorable booklets were returned, and the following year the sample was extended by repeating the procedures for the classes of '39, '44, '49, '54, and '59. In all, a total of 137 completed booklets were scored and the results are given here (see Footnote 3).

The coder in scoring need Achievement following Atkinson manual had no clues in evidence as to any other variables for that subject as the scoring was done. Later the total scores for six pictures and scores for the male and female pictures were summed for each subject.

The test booklet contained a calendar page devised for this study, from which a career pattern could be easily designated roughly according to Super's (1957) categories. Thus one could determine whether a career pattern was uninterrupted without marriage, uninterrupted with marriage ("double track"), interrupted and later resumed, unstable, interrupted and unresumed, or whether there was no career pattern at all.

FINDINGS: THE ACHIEVEMENT MOTIVE IN YEARS AFTER COLLEGE

A comparison of the achievement motive of the five groups in the sample is given in Table 1, a contingency analysis showing the frequencies of high scores compared with those in the medium and low range combined. The scores are obtained from the totals of all six pictures. Each group is composed of members of two successive classes. In these analyses of the Radcliffe data, the cutting point or operational definition of high n Achievement was determined by the nature

of the frequency distribution in the first sample of classes.

The probability that n Achievement and the class membership (or years out of college) are independent is less than .01, and such differences between classes are accepted as significant. The main contributions to chi-square are from the samples 5 and 10 years out of college. The group 5 years out of college has a disproportionate number of women with high n Achievement scores, but in the sample 10 years after graduation a striking decline in achievement motive occurs. Considering just the classes 5 and 10 years after graduation, the probability is less than .001 that such a distribution would occur by chance ($\chi^2 = 10.94$, $df = 1$). From these data it can be inferred that a drop in n Achievement occurs in many of these alumnae between 5 and 10 years after graduation.

There are no significant differences between the groups 15 and 20 years after graduation or between the groups 20 and 25 years later, and the distribution remains stable 15 or more years out of college. Grouping together these latter groups and comparing them with the sample 10 years after graduation the probability is less than .05 that the observed recovery of achievement motive in the sample 15 or more years out of college occurs by chance ($\chi^2 = 5.20$ corrected for continuity, $df = 1$).

From Hypothesis I, one expects this observed decline in n Achievement to correspond

the phase of life when home and family are the major concern and the family is getting established. From Table 2 one can see how the involvement in family and how family structure change by examining the change in family size in these years after college.

The number of children reaches a peak for the group 15 years out of college. The family size tends to remain stable beyond 15 years after college.

Further analysis of the age of the youngest child for these samples of alumnae (Baruch, 1966, p. 104) reveals that there are very few new babies for those 15 or more years beyond college, and the peak number of children under 8 years of age occurs in the group 15 years out of college ($\chi^2 = 67.39$, $df = 24$, $p < .001$). Beginning with the group 15 years out of college, women may still be occupied with child rearing, but more of them become preoccupied again in fantasy with achievement striving.

CAREER PATTERNS AND ACHIEVEMENT MOTIVATION

After the nuclear family has been established, the achievement motive moratorium seems to come to an end. Those women with the highest achievement motive, according to the second hypothesis, are more likely to be found among the employed in these years after the family has been established.

Consider the evidence from the groups 20 and 25 years after graduation with regard to

TABLE 2
NUMBER OF OFFSPRING IN SAMPLES
OF RADCLIFFE ALUMNAE*

No. children	Yrs. out of college					Totals
	5	10	15	20	25	
0	8	5	3	6	6	28
1	4	1	0	1	2	8
2	4	6	3	2	0	15
3	0	3	6	6	3	18
4	0	1	5	2	1	9
5	1	0	2	0	0	3
6	0	0	0	0	1	1
N	17	16	19	17	13	82

Note.— $\chi^2 = 38.1$, $df = 24$, $p < .05$.

* This analysis uses the data from the classes of '38, '43, '48, '53, and '58 and samples an N of 82. There is no evidence to suggest the subsequent year classes were in any way different in family structure.

TABLE 3

CAREER PATTERN AND ACHIEVEMENT MOTIVE
20 OR MORE YEARS OUT OF COLLEGE
(MARRIED ONLY)

n Achievement and career pattern categories	Career pattern		
	Strong	Weak	Total
High	15	4	19
Medium and low	2	25	27
	17	29	46

Note.— $\chi^2 = 21.523$ corrected for continuity, $df = 1$, $p < .001$.

this hypothesis (although in this sample only 48% of this group is in paid employment). A dichotomy is established such that strong career patterns include the uninterrupted "double track" and the "interrupted and later resumed" patterns, and weak career patterns are comprised of the interrupted and unresumed career pattern, the unstable, and the completely absent patterns (after Super's, 1957, categories). Using the total score for need achievement, no significant association between n Achievement and career pattern was found. Remembering the findings of Lesser et al. that "achieving" girls score significantly higher on the picture with female central figures, only the subscores for the three female pictures were used in the analyses of Table 3. The association between n Achievement as projected on female central figures and career pattern is high for the married women 20 and 25 years out of college ($p < .001$). Married women with a high n Achievement who are 20 or more years beyond college are very likely to be found pursuing careers, whereas others with low and medium levels of n Achievement are not likely to be found so employed.

For the group 15 years out of college, where some resurgence of achievement motivation was found but also there is still the presence of young children in the home, no significant association between n Achievement and career pattern exists ($.10 < p < .20$), and there are in this group women with high achievement motivation but weak (unresumed) career patterns. Interpreting these data as part of a developmental sequence, one would predict that some of the women

in the group 15 years out of college with high achievement motivation will resume careers within the next 5 years. The high achievement motivation is conceived here as the antecedent variable and the career behavior or pattern as the consequent variable.

There are other pieces of evidence which support the findings reported here. A survey (United States Department of Labor, 1962) by the Women's Bureau finds the group 15 years out of college reporting dissatisfaction with the limitations of women's role activities and eager to resume careers. Friedan's (1963, p. 27) study of 200 of her classmates 15 years after their graduation from Smith College also reports the yearning of women at this stage for "something more than" husband, children, and home. Roe (1966) cites a scientist who makes a similar observation of his former women students who go through a period of low level of enterprise while raising a family, but "tend to come back."

These data alone cannot establish whether the interpretation of a developmental cycle is justified, and the pattern attributed here to a temporal cycle may alternatively be attributed to differences in generations of women. A longitudinal study or replication of this study at a later period can help establish one interpretation over the other.

A BROADER TEST OF THE HYPOTHESES

The hypotheses, confirmed by the data from a very selected sample of college alumnae, were then tested with data from a nationally representative sample of 763 women who had participated in a study conducted by the Survey Research Center (Gurin et al., 1960). The multiple contingency analyses of these data, discussed more fully elsewhere (Baruch, 1966), allowed simultaneous consideration of four variables: work status, education, Achievement, and age. Neither hypothesis could be supported by these data. Although achievement motivation and age were found to be associated, the pattern for the population in general was not as predicted. Instead, the major effect observed was the decline of achievement motivation in the groups over 55. Questions may arise about the appropriateness of the age of the central figures for these older subjects. For the col-

lege-educated women in this national sample, however, the temporal pattern resembled that which did not accurately duplicate, that pattern found in the Radcliffe sample. These college women were not necessarily college graduates, but may have had, by self-report, as little as 1 year of college. The pattern of resurgence of achievement motive was found for these college women, with the age group from 35-39 most likely to have high achievement motivation. Different patterns of achievement motive with age for women at different educational levels were statistically significant at the .02 level, as indicated by the interaction chi-square among these three variables. For the women with a high school or equivalent education, those in the youngest age group (21-24) scored high, with a decline in the 30-39 age group and some subsequent rise. For those women with less than 12 years of education, there was a steady decline in the incidence of high achievement motivation with age.

Temporal patterns of achievement motive then, appear to differ for women of different educational backgrounds. The relation specified in the first hypothesis appears to hold only for a subset of the total population. If we assume that the motivation measured is a resultant of arousal cues present in the environment as well as internal motivation, then the finding seems reasonable that Achievement scores will be highest for those closest to the school years. For those poorly educated women who are less likely to be in a stimulating environment, the decline is continuous. For the college-educated women, who, we presume, are living more stimulating lives, arousal cues will be present to heighten, once again, the level of achievement motivation, returning this motive to regnancy.

The second hypothesis must be rejected on the basis of this test, for work status and achievement motive are independent when one considers the population in general. For those women who have attained only a high school education or less than a high school education, economic factors may be more important determinants of work status than achievement striving or personal satisfaction. The interaction (chi-square) among work status, education, and age is significant, indicating that the pattern of employment with

different for women of different educational backgrounds.

The achievement motive cannot be said to account for the employment patterns, for no significant association is found. However, upon examining work status and achievement motive simultaneously as a function of age, for women of each educational level, a pattern emerges suggesting that the increase in achievement motive, when it occurs, is followed by a subsequent return to paid employment some years later. A cross-correlation function (Goldman, 1953, pp. 237-238), when computed for the college subpopulation, indicates a time lead of achievement motive with work status following (Baruch, 1966, pp. 161, 163). This last inference, quite speculative and based on a hypothesis formulated after these data were examined, should be investigated further.

The college group, for whom this trend is most visible, are the women who are most likely to have a freer choice, economically, about working or not working. Further, they are more likely to be qualified to hold jobs which can satisfy the achievement motive. The opportunities are greater for the college-educated women to move from a preoccupation with achievement in fantasy to a later expression of this need through vocational behavior.

REFERENCES

- Alumnae Directory of Radcliffe College*, 1962, 27, No. 4.
- ATKINSON, J. W. (Ed.) *Motives in fantasy, action, and society*. Princeton, N. J.: Van Nostrand, 1958.
- BARUCH, R. The achievement motive in women: A study of the implications for career development. Unpublished doctoral dissertation, Harvard University, Graduate School of Education, 1966.
- CHILD, I. L., STORM, T., & VEROFF, J. Achievement themes in folk tales related to socialization practice. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton, N. J.: Van Nostrand, 1958. Pp. 479-492.
- DEGLER, C. N. Revolution without ideology: The changing place of women in America. *Daedalus*, 1964, 93, 653-670.
- DEUTSCH, H. *The psychology of women*. New York: Greene & Stratton, 1944.
- FLEMING, T. J. The world of women at work. *Boston Herald Magazine*, February 9, 1964.
- FRENCH, E., & LESSER, G. S. Some characteristics of the achievement motive in women. *Journal of Abnormal and Social Psychology*, 1964, 68, 119-128.
- FRIEDMAN, B. *The feminine mystique*. New York: Dell, 1963.
- GOODMAN, L. The changing pattern of women's work. *Social Psychology Quarterly*, 1965, 28, 11-21.
- GOODMAN, S. *Information theory*. Englewood Cliffs, N. J.: Prentice Hall, 1961.
- GOLDENFAUL, D. (Ed.) *Information please almanac, atlas and yearbook*. New York: Simon & Schuster, 1964.
- GURIN, G., VEROFF, J., & FIELD, S. *Americans view their mental health*. New York: Basic Books, 1960.
- LESSER, G. S., KRAWITZ, R. N., & PACKARD, R. Experimental studies of achievement motivation in adolescent girls. *Journal of Abnormal and Social Psychology*, 1965, 66, 1-14.
- MATTHEWS, E. The marriage career conflict in girls and women. Unpublished doctoral dissertation, Harvard University, Graduate School of Education, 1966.
- MCCHESLAND, D. C. *The achieving society*. Princeton, N. J.: Van Nostrand, 1961.
- NATIONAL MANTOWER COUNCIL. *Womanpower: Work in the lives of married women*. New York: Columbia University Press, 1965.
- NYE, F. I., & HUTTMAN, L. W. The socio-cultural setting. In F. I. Nye & L. W. Huttman (Eds.), *The employed mother in America*. Chicago: Rand McNally, 1965. Pp. 3-17.
- PARSONS, T., & BALES, R. F. *Family, socialization, and interaction process*. Glencoe, Ill.: Free Press, 1955.
- PRESIDENT'S COMMISSION ON THE STATUS OF WOMEN. *American women*. Washington, D. C.: United States Government Printing Office, 1960.
- ROE, A. Satisfaction in work. *Journal of the American Association of University Women*, 1961, 55, 230-233.
- ROE, A. Women in science. *Personnel and Guidance Journal*, 1960, 54, 784-787.
- ROSENFELD, C., & PERRELLA, V. C. Why women start and stop working: A study in mobility. *Monthly Labor Review*, September 1965 (Reprint No. 2476).
- SUPER, D. E. *The psychology of careers*. New York: Harper, 1957.
- UNITED STATES DEPARTMENT OF LABOR, Women's Bureau. *15 years after college: A study of alumnae of the class of 1945*. (Bulletin 283) Washington, D. C.: United States Government Printing Office, 1960.
- VEROFF, J., ATKINSON, J. W., FIELD, S. C., & GURIN, G. The use of thematic apperception to assess motivation on a nationwide interview study. *Psychological Monographs*, 1960, 74(12, Whole No. 499).
- VEROFF, J., WILCOX, S., & ATKINSON, J. W. The achievement motive in high school and college age women. *Journal of Abnormal and Social Psychology*, 1953, 48, 108-119.

(Received April 15, 1966)

SELF-ESTEEM AND SUSCEPTIBILITY TO SOCIAL INFLUENCE¹

RICHARD E. NISBETT² AND ANDREW GORDON

Department of Social Psychology, Columbia University

It was hypothesized that the form of the relationship between self-esteem and influence is dependent on the comprehensibility and plausibility of the influence induction. Negative relationships will be found when inductions are relatively easy to comprehend or relatively implausible. More difficult or plausible inductions will produce nonmonotonic or even positive relationships. Consistent with these notions, a nonmonotonic relationship was found in the present study when the induction was relatively easy to comprehend, but not highly plausible, and a slightly positive relationship was found when the induction was difficult to comprehend, but highly plausible. Interactions between manipulated and predispositional self-esteem which were obtained were opposite to those predicted. This result was discussed in terms of Cohen's ego-defense notions.

A number of students of persuasion have reported a monotonic negative relation between self-esteem and susceptibility to social influence (Asch, 1958; Berkowitz & Lundy, 1957; Janis, 1954, 1955; and several studies reported in Janis & Hovland, 1959). Recent work has, however, suggested that the relationship between these two variables may be nonmonotonic (Cox & Bauer, 1964; Gergen & Bauer³). Indeed, a study reported by Appley and Moeller (1963) shows that 33 of 38 personality variables were nonmonotonically related to persuasibility, even within a rather homogeneous population.

McGuire⁴ (1966, in press) has suggested a multiplicative two-factor model to account for these nonmonotonic relations between personality variables and susceptibility to social influence. Briefly, it is postulated that opin-

ion change is the outcome of a chain of processes including (a) attention and comprehension of the persuasive message, and (b) yielding to what is comprehended of the position in the message. To predict how a personality variable is related to comprehension, one must consider two things: first, how the personality variable is related to comprehension and to yielding, and second, how much individual difference variance there is in comprehension and in yielding in the given influence situation.

The nonmonotonic case arises when the given personality variable is related to comprehension and to yielding in opposite directions. The quantitative considerations are presented in McGuire (1961; see also Footnote 4) and in Bogartz (1965). McGuire (in press) has argued that self-esteem is such a variable. Since self-esteem is likely to be positively related to intelligence and interest in the outside world, etc., it will be positively related to the amount of attention to and accurate comprehension of the persuasive message. Since self-esteem is likely to be negatively related to feelings of inadequacy and lack of ego strength, it is likely to be negatively related to yielding. Hence, insofar as yielding is concerned, high self-esteem would protect the person from persuasion. According to the quantitative models in the above-mentioned papers, the maximum persuasibility should occur at some intermediate level of self-

¹ The authors wish to acknowledge valuable contributions made by William J. McGuire to all phases of this study. We are also grateful to Stanley Lehmann and to Joel Savell for their valuable assistance. The research was supported by Grant G-19799 from the Division of Social Sciences, National Science Foundation, to McGuire.

² Now at Yale University.

³ K. J. Gergen and R. A. Bauer, "The Interactive Effects of Self-Esteem and Task Difficulty on Social Conformity," unpublished manuscript, Harvard University, 1966. We are indebted to K. J. Gergen for making his manuscript available and for a critical reading of our own.

⁴ W. J. McGuire, "Effectiveness of Fear Appeals in Advertising," Advertising Research Foundation, August 20, 1963. (Mimeo)

where the comprehension and yielding gradients cross.

According to the model, this optimal level of self-esteem for producing susceptibility to social influence moves up or down depending on the type of influence induction. Consequently, the relationship between the two variables can range from monotonically negative to nonmonotonic to monotonically positive, depending on the characteristics of the influence situation. McGuire (in press) has argued that the more difficult the reception (attention and comprehension) aspects of the persuasion situation, the higher will be the level of self-esteem at which maximum persuasion occurs. The more difficult (due to lack of evidence or implausibility or whatever) is yielding to the message, the lower will be the level of self-esteem at which maximum persuasion occurs.

The present study tested the implications of this model by varying both the reception and yielding properties of persuasive messages. Two influence inductions were employed. The first was easy to receive but difficult to yield to, consisting of mere mention of a purported position taken on an issue by an authoritative source. The second employed material which was difficult to receive but very easy to yield to. This induction was a rationally argued communication employing factual-seeming arguments. Both the reception and the yielding properties of the first induction are geared toward producing maximum influence for relatively low self-esteem subjects, while the reception and yielding properties of the second induction favor influencibility for relatively high self-esteem subjects. An interaction between level of self-esteem and method of influence induction is therefore predicted, such that maximum opinion change should occur at a lower level of self-esteem for the easy to receive, difficult to yield to induction than for the difficult to receive, easy to yield to induction.

A second aspect of McGuire's (1966, in press) theorizing about personality-persuasibility relations is also tested by the present study. The main point of his Principle 4 (McGuire, in press) is that we cannot equate acute, manipulated changes of personality variables with chronic, predispositional levels

of the personality variables in predicting personality-persuasibility relations. However, he suggests as a tactic for detecting non-monotonic relations that the acute and chronic levels can be used additively. An example of the additivity of acute and chronic personality variables for the case of anxiety is provided by McGuire (see Footnote 4), Millman (1965), and other studies reviewed by McGuire (1966). When a person with low chronic anxiety is made more anxious by an acute manipulation, he becomes more persuasible, while, when anxiety is acutely added to the state of an individual with already high chronic anxiety, he becomes less persuasible. McGuire (in press) has suggested that the same may be true for self-esteem, and Gelfand's (1962) results with verbal conditioning data tend to support this notion. To test this notion, an acute manipulation of self-esteem in the form of a success or failure induction is included in the present design. It is predicted that, for both influence attempts employed, there will be an interaction between chronic and acute self-esteem such that a failure induction will be relatively more effective with subjects chronically high in self-esteem and a success induction will be relatively more effective with subjects chronically low in self-esteem.

METHOD

General Procedure

The experiment was conducted in two sessions, 2 weeks apart. Under the guise of helping the state department of education to standardize test materials, subjects completed two self-esteem tests and an intelligence test at the first session. At the second session, subjects received the "results" of the intelligence test, which indicated to each subject that he had done extremely well or extremely poorly. Subjects then completed both the self-esteem tests for the second time and read various persuasive communications. Finally, subjects indicated their opinions on issues dealt with in the communications they had read. Debriefing followed immediately.

Personality Materials

Acute self-esteem manipulation. All subjects were given a purported "intelligence test" consisting of items resembling those on standard intelligence tests, plus some impossible and very ambiguous items which were included so that subjects would find it plausible that they had done very well or very

poorly on the test. In the final session half the subjects were informed that they had failed (scored in the fifteenth, sixteenth, or seventeenth percentile—the manipulated low-self-esteem condition), and half that they had succeeded (scored in the ninety-fifth, ninety-sixth, or ninety-seventh percentile—the manipulated high-self-esteem condition). The percentiles were purportedly based on norms for New York college students. In order to get the point, and as a check on the manipulation, those scoring below 20 or above 80 (i.e., all subjects) were asked to indicate if any special circumstances were responsible for their low or high scores. If there were no such circumstances, they were to write "none."

Chronic self-esteem measures. The measures of preexperimental self-esteem were presented to the subjects as "insight tests." The first of these measures consisted of 44 statements dealing with feelings of general self-confidence culled largely from the MMPI and reworded so as to be more suitable for college-age normals. The items used can be found in Conlon (1965). The scale is similar to Janis and Field's (1959) Feelings of Inadequacy Scale and includes some of the same items reworded for scale format. Subjects were asked to indicate their personal feelings about each statement (e.g., "I am a pretty fast learner"; "I have frequently given up doing a thing because I thought too little of my ability") on a 7-point scale, with the ends labeled "Definitely False" and "Definitely True." Items were counterbalanced for acquiescence. The second measure was a list of polar evaluative adjectives in semantic differential format (Fiedler, 1958) to which subjects responded by describing "Myself as I Am" and "My Ideal Self." The results are here reported in terms of the self-esteem scores obtained on the MMPI-type questionnaire, but the conclusions would have been quite similar had the Myself as I Am score or the self-ideal discrepancy score been used as the measure of chronic self-esteem.

Persuasibility Materials

Issues. One month prior to the first session, one of the three introductory psychology classes used in the study was administered a questionnaire consisting of various statements about medical concepts. Included in the list of medical statements were several medical truisms of the type McGuire (1961) described. (E.g., "Everyone should brush his teeth after every meal if at all possible.") Subjects were asked to indicate the extent of their agreement with each of the statements on a 15-point scale. The results of this premeasure were used to assign truisms of equal mean acceptance to each of the attack and control conditions. Mean acceptance of the truisms employed in the study was quite high, averaging 12.84 out of a possible 15.

Unsubstantiated attack. As a favor to the state health department, subjects were asked to read a list of "health facts," allegedly compiled by the state

medical association. Subjects were then to indicate whether they knew each fact and whether they thought their friends knew it. This task was included to insure their careful reading of each item. 10 of the 15 "health facts" were reasonable. "Mouth-to-mouth resuscitation is the most effective means of artificial respiration." However, presented information which the premeasure had shown to be contrary to the beliefs of the subjects: "Provided he is in good health, a person can donate blood to any other person regardless of blood type"; "Since the incidence of the childhood diseases diminishes with age, it is not necessary to contract them as an adult." These statements constituted the unsubstantiated attack items. No influence attempt was made beyond the single exposure allegedly stating the medical association's views.

Substantiated attack. As a further service to the state health department, subjects were asked to read and evaluate two communications to be used in an forthcoming health campaign. The messages, 650 words long, were titled "Some Harmful Effects of Chest X-rays" and "Some Dangers of Frequent Tooth Brushing." Seemingly well-grounded in scientific fact and presented in a logical manner, these communications typically produce considerable weakening of belief in the advisability of an annual chest X-ray and regular tooth brushing (McGuire, 1961). The truisms which these messages attacked were equated with the unsubstantiated attack truisms in terms of initial belief level on the premeasure. Half of the subjects received the substantiated attack first, and half received the unsubstantiated attack first.

Unsubstantiated support items and control items. As mentioned above, 13 of the 15 "health facts" purportedly endorsed by the state medical association did indeed coincide with the subjects' initial perceptions. These were included to add credibility to the unsubstantiated attack. Two of these supported truisms were equated for initial belief level with truisms employed in the attack methods. It is important to have an annual medical checkup; and "Moving an injured person is usually more dangerous than simply covering him with a blanket and waiting for help to arrive." These statements provided the unsubstantiated support data discussed below. An additional four truisms, equated with the two unsubstantiated support truisms and the two attack truisms as regards initial levels, served as controls. These were neither attacked nor defended and were not referred to until the final questionnaire.

Persuasibility measure. A final Medical Opinions Questionnaire presented 30 health statements with which subjects indicated their agreement on a 15-point scale (McGuire, 1961). These included two items, counterbalanced for acquiescence, for each of the persuasive issues discussed above (two substantiated attack, two unsubstantiated attack, and two unsubstantiated support) and four control items, also counterbalanced.

Design and subjects. Three levels of chronic self-

were distinguished on the basis of responses to the MMPI-type questionnaire. Since the success-failure manipulation was given randomly to approximately half of high, medium, and low self-esteem groups were high on acute self-esteem and half were low. Every individual in the high and low groups read both the unsubstantiated and substantiated attacks, half in that order and half in the reverse order. Subjects were 152 students in three introductory psychology courses at Long Island University.

RESULTS

Effect of Success-Failure Manipulation on Chronic Self-Esteem Measures

The check on the subject's perception of the manipulation indicated that at least 88% of the subjects perceived correctly, in accord with the success or failure induction, that they had done extremely well or extremely poorly. Awareness did not vary as a function of chronic self-esteem. Since the results are not altered by omitting subjects who failed to demonstrate their understanding of the manipulation, their data are retained in the analyses reported below.

Surprisingly, we failed to detect any sizable effect of the success-failure manipulation on the measures of self-esteem. Although the successful subjects have higher scores on each of the second-administration self-esteem measures than the failure subjects, none of the differences attains conventional levels of statistical significance. Controlling for initial level of self-esteem, the F ratios of the MMPI-type questionnaire, Myself as I Am scores, and self-ideal discrepancy scores are, respectively, 2.49, .71, and 2.61 ($p = .12$, .40, .11, respectively; $df = 1/146$).^a

This failure of the difference to attain significance is surprising because the relief or disappointment of the subjects upon learning the bogus nature of their IQ scores was both audible and visible to the experimenters. Since the self-esteem measures were given immediately after subjects received their IQ scores, it may be that this interval was too short to allow self-esteem effects to seep down to the test response. It may also be that the paper-and-pencil tests are insensitive to temporary changes in self-esteem.

^a All p values reported in this paper are two-tailed.

TABLE 1
MEAN REJECTION OF TRUISMS BY THE
UNSUBSTANTIATED ATTACK METHOD

Acute self-esteem	Chronic self-esteem	Mean	SD	N
Substantiated	High	27.78	10.25	25.04
A	Medium	20	28	75
Unsubstantiated	High	23.91	10.17	17.86
A	Medium	11	22	77
Substantiated	Low	22.43	11.04	21.88
A	Medium	11	40	152

Effect of the Variables on Influencibility

Mean influencibility as a function of chronic and acute self-esteem is presented for the unsubstantiated attack in Table 1 and for the substantiated attack in Table 2.^a Scores are based on the extent to which subjects reject the attacked truisms and are sums of 4 15-point scales. Thus a score of 4 would indicate complete acceptance of the truisms (and total absence of influence) and a score of 60 would indicate complete rejection of the truisms (and maximum influencibility). As would be expected, the substantiated attack was considerably more effective in producing opinion change than the unsubstantiated attack ($p < .001$).

Effects of the chronic self-esteem variable. It had been predicted that there would be an interaction between level of chronic self-esteem and type of influence induction. The unsubstantiated attack, since it is easily comprehended and not very plausible, will be most effective with subjects having a relatively low level of self-esteem. The substantiated attack, since it is harder to comprehend and more plausible, will reach its peak effectiveness at a relatively higher level of self-esteem.

This prediction is well supported by the results. It can be seen in Tables 1 and 2 that the peak influencibility does indeed occur at a lower level of chronic self-esteem for the unsubstantiated attack than for the substantiated attack. Peak influencibility is achieved at the medium self-esteem level for the unsubstantiated attack and at the high

^a The tables are not partitioned on the basis of order. This variable was manipulated only as a control, not because it was of interest in itself.

TABLE 2
MEAN REJECTION OF TRUISMS BY THE
SUBSTANTIATED ATTACK METHOD

Attack method	Chronic self-esteem			
	Low	Medium	High	Total
Sum of scores	13.78	36.73	42.43	92.94
N	21	20	20	61
Percentage	33.71	60.84	41.14	60.57
N	24	22	22	68
Total	37.97	58.94	64.66	161.57
N	51	51	50	152

self-esteem level for the substantiated attack. The interaction between chronic self-esteem and attack method is significant at the .05 level ($F = 3.20$, $df = 2/146$).

The form of the relationship between chronic self-esteem and influencibility may be profitably examined in detail. The main effect of chronic self-esteem for the unsubstantiated attack method departs significantly from a horizontal line ($F = 3.71$, $df = 2/146$, $p < .05$). It may be seen in Figure 1 that the relationship between self-esteem and influencibility for this attack method is a markedly curvilinear one. The curve is based on the sum of the two issues, but the separate issue profiles have almost identical shapes. The present model leads us to expect that, except for specifiable communication situations, this is the characteristic form of the relationship.

Given that the relationship between chronic self-esteem and influencibility is nonmono-

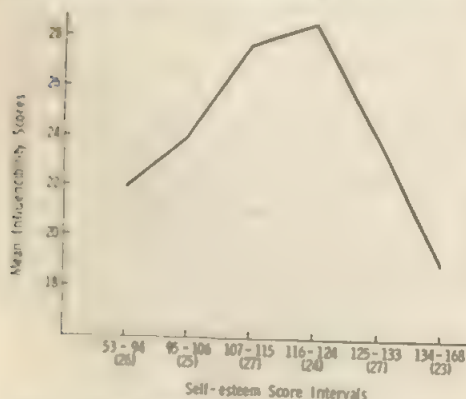


FIG. 1. Influencibility by the unsubstantiated attack method as a function of self-esteem. (Figures in parentheses denote N for the cell.)

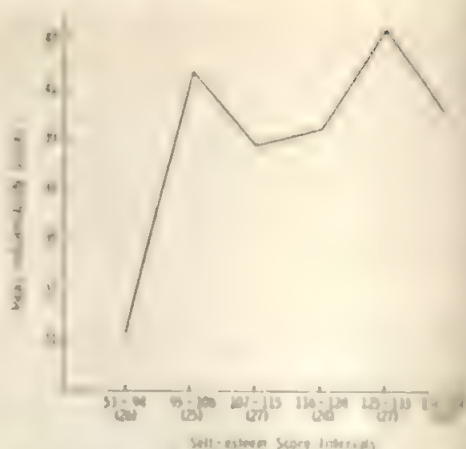


FIG. 2. Influencibility by the substantiated attack method as a function of self-esteem. (Figures in parentheses denote N for the cell.)

tonic with the unsubstantiated attack, the model leads us to expect one of two relationships with the substantiated attack. Either the influencibility curve should begin to fall at some relatively high level of self-esteem or, under the assumption that the receptivity and yielding curves do not cross at all for this influence induction, the relationship should be a positive monotonic one. Figure 2 shows the form of the relationship between self-esteem and influencibility by the substantiated attack method. The curve shows a generally positive relationship between the two variables, and this is true also of the separate issue profiles. While a drop does occur between the fifth and sixth sextiles, the difference is statistically trivial. Apparently the receptivity and yielding curves do not therefore, cross at all for this influence induction. It would be convenient then to say that there is an unequivocally positive relationship between self-esteem and influencibility by the substantiated attack method. This cannot be done without hesitation, however, since the correlation between the two is only .15 and of borderline significance ($p = .07$, $df = 150$). At any rate, the form of this relationship is quite different from the negative relationships found in the past and from the curvilinear one in the present study.

The results with regard to the main concerns are clear. Maximum influencibility

ages at a lower level of self-esteem for the case of the unsubstantiated, implausible induction than for the difficult, plausible induction and for at least one of the inductions an unequivocally curvilinear relationship emerges. These results are quite similar to those reported by Gergen and Bauer (see Footnote 3). With the less difficult inductions they obtained a curvilinear relationship between self-esteem and influencibility. With their more difficult induction, a trivially positive relationship was obtained. Since these authors speculate that their findings may be unique to female populations, it should be pointed out that the results in this study were entirely similar for subjects of both sexes.

The finding of a nonmonotonic relationship between self-esteem and influencibility in the case of the unsubstantiated attack and the finding of a slight positive monotonic trend in the case of the substantiated attack appear anomalous when viewed from the context of many of the prior studies of self-esteem and persuasibility cited in our introductory paragraph. The generalization drawn from these studies that there is a monotonic negative relationship between the variables is called into question by the present findings and model. According to McGuire's (in press) theory, we would expect to find a negative relationship only to the extent that some or all of the following were true: the communication concerns matters of opinion rather than "fact"; the communicator is less than completely expert and trustworthy; the arguments are irrational or nonexistent; or reception of the message is total for individuals at all self-esteem levels. Similarly, if one wanted to obtain an unequivocally positive relationship between self-esteem and influencibility, one would simply invert all these communication characteristics. Past studies which have found negative relationships have typically used persuasion materials which were relatively difficult to yield to. For instance, the studies by Janis and his colleagues used issues which were matters of opinion, rather than fact: "Hindenburg Called German Hero, Democratic Leader." The controversiality of the issues was in fact stressed in the instructions given to subjects in these studies. Similarly, the communicators in the studies were

less than completely authoritative. In contrast studies finding a nonmonotonic relationship have typically used issues which were relatively factual or uncontroversial, and more than fully authoritative or both. The present study also employed attacks which were easy to yield to. This was of course done deliberately in order to maximize the likelihood of obtaining nonnegative relationships. Thus the communication in the present study was a parody of phenomena of the state medical association, and the issues were questions of scientific fact. The substantiated attack, which added detailed, convincing argument to these communications, effectively produced a relationship between self-esteem and influencibility which is the antithesis of that found in earlier studies. *

Effects involving the acute self-esteem variable. It had been predicted that with both influence inductions there would be an interaction between acute and chronic self-esteem. Subjects with chronically high self-esteem should be made relatively more influencible by a failure induction; subjects with chronically low self-esteem should be made relatively more influencible by a success induction.

As can be seen in Tables 1 and 2 there was no significant main effect of manipulated self-esteem with either method, nor was one predicted, since the hypothesis predicted rather an interaction between acute and chronic self-esteem for both methods. In both Tables 1 and 2, subjects with chronically low self-esteem should have been more influencible after receiving a success experience, and subjects with chronically high self-esteem should have been more influencible after a failure experience. With neither attack method was the predicted interaction found, and in fact it can be seen that with both attack methods the findings are precisely opposite to the predictions, though these trends did not attain conventional levels of statistical significance. The data in Tables 1 and 2 show that in both cases low-self-esteem subjects are, if anything, more influencible after experiencing failure, while high-self-esteem subjects are more influencible after experiencing success.

Certainly the prediction was not confirmed,

TABLE 3
MEAN REJECTION OF TRUISMS BY THE
UNSUBSTANTIATED SUPPORT METHOD

Acute self-esteem	Chronic self-esteem			
	Low	Medium	High	Total
Success	13.41	13.80	9.61	12.09
N	27	20	28	75
Failure	10.83	11.74	14.50	13.41
N	24	31	22	77
Total	12.20	12.55	11.76	12.76
N	51	51	50	152

but whether the actual form of the interaction is the reverse of that predicted, as suggested in Tables 1 and 2, is left in doubt by the lack of statistical significance. Fortunately, there are other data within the present design where the reversal of the prediction attains significance. It will be recalled that the list of "health facts" containing the unsubstantiated attack items also included unsubstantiated supports for two additional health beliefs. These items, which supported the subject's health preconceptions, were included to add plausibility to the attacks on the other preconceptions, but in effect these supports also constitute an influence induction in that they urge subjects to believe still further in the health truisms. With these unsubstantiated support items, precisely the same interaction occurs as occurred with the attack methods—a reversal of our original interaction prediction. With these unsubstantiated support items, the interaction becomes highly significant ($F = 5.06$, $df = 2/146$, $p < .01$). Mean influencibility scores for the unsubstantiated support items are presented in Table 3.⁷ (It should be noted that, unlike Tables 1 and 2, influencibility in this table is associated with *low* rejection of truisms.) The model predicts that for subjects with high self-esteem a failure experience should result in greater influencibility (less rejection of the supported truisms) than a success experience. It can be seen that the opposite was found. Similarly, the model predicts that a success experience should result in less rejection of

the truisms for subjects with low self-esteem. Again, the opposite was found. (With the four unmentioned control items, where there was no influence attempt, no hint of the interaction was found.)

DISCUSSION

Interpretation of the chronic-acute interaction. Reasoning from Cohen's (1939) theory that different ego defenses are employed by individuals with high and low self-esteem, Silverman (1964a, 1964b) has predicted and found a similar interaction. Briefly, Cohen's theory states that individuals with high self-esteem maintain their confidence by avoidance defenses, whereas the characteristic defenses of low-self-esteem individuals are expressive—for example, projection. A high-self-esteem individual will tend to avoid unfavorable information about the self altogether, while a low-self-esteem individual distorts it. Moreover, the low-self-esteem individual tends to respond to failure by identifying with others, under the assumption that their beliefs and actions are more successful in providing need satisfaction than his own. All of this leads to the prediction that high-self-esteem individuals should be made less influencible by a failure experience, due to their tendency to use avoidance defenses, while the reverse should be true of subjects with low self-esteem.

Silverman's results are in line with this interpretation and so are the present findings. With both of the attack methods high-self-esteem subjects tend to be less influencible after a failure manipulation, while low-self-esteem subjects tend to be more so. When the persuasive effort is in the direction of bolstering belief in truisms—with the unsubstantiated support items—this interaction becomes highly significant. The Cohen-Silverman hypothesis cannot account for this fact that the interaction is stronger when the truisms are bolstered than when they are attacked, but it is otherwise quite consistent with our data. Perhaps, as Silverman (1964a) suggests, the different results in Gelfand's (1962) study are due to her using young subjects whose defensive styles had not yet differentiated clearly.

⁷ Because of the greatly lower variance of the scores on this variable, mean differences of considerably less magnitude than those in Tables 1 and 2 are considerably more significant.

Conclusions. The general assumption of McGuire's (in press) theory—that the mediating role of reception, as well as that of yielding, must be considered in predicting personality-persuasibility relationships—was well supported by the present study. The heuristic power of this principle was exhibited in various aspects of our findings: the nonmonotonicity of the relationship between chronic self-esteem and influencibility by unsubstantiated (easily comprehended) attacks, the tendency for higher levels of chronic self-esteem to be optimal for influencibility as the persuasive attack becomes more difficult to comprehend, and the necessity to consider avoidance and distortions of reception to account for chronic-acute interactions.

On the other hand, the results show that the auxiliary hypothesis of simple additivity of chronic and acute variation of a given personality dimension, which McGuire (in press) makes in connection with his Principle 4 on "confounding," is quite oversimplified. While it sometimes predicts the obtained results (Gelfand, 1962; McGuire, 1966; Millman, 1965), the results can at other times not only fail to confirm it, but even be significantly in the opposite direction, as in this study and the studies of Silverman (1964a, 1964b). In making the simple additive hypothesis, in connection with his Principle 4, McGuire (1966) neglected the main point of this confounding principle: that when we use variations in the person's chronic level on a given personality characteristic, we must consider that the trait will be embedded in a set of confounding characteristics that the person has acquired to protect himself from the maladaptive tendencies inherent in the initial characteristic. In this study our data suggest the nature of some of this confounding in the case of chronic self-esteem and show that an interaction between chronic and acute self-esteem and influencibility can be quite different from McGuire's simplistic additive hypothesis. Though chronic and acute states of anxiety interacted "additively" in Millman's (1965) study, our results indicate that chronic and acute self-esteem do not relate in this way. This unfortunate conclusion (from the standpoint of compact theory in the area of personality and influencibility) is made

more likely by the fact that the success-failure manipulation had little impact on the three measures of self-esteem. Taken together, the results suggest that the effect of a success-failure manipulation is not to alter self-esteem, but to trigger ego-defense mechanisms which differ for individuals at different self-esteem levels.

The basic notions of the model—that chronic self-esteem is related in opposite directions to reception and yielding and that the form of the self-esteem-influencibility relationship is determined by characteristics of the communication—are well supported by the present study.

REFERENCES

- APPLEY, M. H., & MOELLER, G. Conforming behavior and personality variables in college women. *Journal of Abnormal and Social Psychology*, 1963, 66, 284-290.
- ASCH, S. E. Effects of group pressure upon the modification and distortion of judgments. In E. E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1958. Pp. 174-182.
- BERKOWITZ, L., & LUNDY, R. M. Personality characteristics related to susceptibility to influence by peers or authority figures. *Journal of Personality*, 1957, 25, 385-397.
- BOGARTZ, R. S. On the assumption of a steeper avoidance gradient in Miller's conflict theory. *Psychological Review*, 1965, 72, 162-163.
- COHEN, A. R. Some implications of self-esteem for social influence. In C. I. Hovland & I. L. Janis (Eds.), *Personality and persuasibility*. New Haven: Yale University Press, 1959. Pp. 102-120.
- CONLON, E. J. Performance as determined by expectation of success or failure. Unpublished doctoral dissertation, Columbia University, 1965.
- COX, D. F., & BAUER, R. A. Self-confidence and persuasibility in women. *Public Opinion Quarterly*, 1964, 28, 453-466.
- FIEDLER, F. E. *Leader attitudes and group effectiveness*. Urbana: University of Illinois Press, 1958.
- GELFAND, D. M. The influence of self-esteem on rate of verbal conditioning and social matching behavior. *Journal of Abnormal and Social Psychology*, 1962, 65, 259-265.
- JANIS, I. L. Personality correlates of susceptibility to persuasion. *Journal of Personality*, 1954, 22, 504-518.
- JANIS, I. L. Anxiety indices related to susceptibility to persuasion. *Journal of Abnormal and Social Psychology*, 1955, 51, 663-667.
- JANIS, I. L., & FIELD, P. B. Sex differences and personality factors related to persuasibility. In I. L. Janis & C. I. Hovland (Eds.), *Personality and persuasibility*. New Haven: Yale University Press, 1959. Pp. 55-68.

- JANIS, I. L., & HOVLAND, C. I. (Eds.) *Personality and persuasibility*. New Haven: Yale University Press, 1959.
- MCGUIRE, W. J. Resistance to persuasion conferred by active and passive prior refutation of the same and alternative counterarguments. *Journal of Abnormal and Social Psychology*, 1961, 63, 326-332.
- MCGUIRE, W. J. Attitudes and opinions. In P. F. Farnsworth (Ed.), *Annual review of psychology*. Vol. 17. Palo Alto, Calif.: Annual Review, 1966. Pp. 475-514.
- MCGUIRE, W. J. Personality and susceptibility to social influence. In E. F. Borgatta & W. W. Lambert (Eds.), *Handbook of personality theory and research*. Chicago: Rand McNally, in press.
- MILLMAN, S. The relationship between anxiety, learning and opinion change. Unpublished doctoral dissertation, Columbia University, 1965.
- SILVERMAN, I. Differential effects of ego threat upon persuasibility for high and low self-esteem subjects. *Journal of Abnormal and Social Psychology*, 1964, 69, 567-572. (a)
- SILVERMAN, I. Self-esteem and differential responsiveness to success and failure. *Journal of Abnormal and Social Psychology*, 1964, 69, 115-119. (b)

(Received April 21, 1966)

A DEVELOPMENTAL EXAMINATION OF PERFORMANCE IN A TACIT COORDINATION GAME SITUATION¹

CHARLES L. FRY

University of Virginia

The tacit coordination game performances of 24 4th graders and 24 8th graders were compared with college controls. Ss played 3 2-person, 3-choice games with several like-aged partners from different ascendance quartiles. The game required Ss to repeatedly anticipate the object choice of a partner and provided increasing information for doing so. College Ss outperformed 8th graders who outperformed 4th graders ($p's < .001$). College and 8th-grade Ss improved with successive partners ($p's < .06, .001$). The pattern and object convergence strategies found in the college Ss were less evident in 8th graders and were absent in 4th graders. Differences between partners' ascendance scores predicted game success in college and 8th-grade Ss only.

In the tacit coordination game (TCG), game participants are isolated from one another and are asked to anticipate their partners' responses. Schelling (1960) outlined some TCG games, and predicted that subjects could anticipate their partners' responses at levels exceeding chance because of their ability to view the game situation from points of view other than their own. If a subject is allowed a series of attempts to anticipate his partner's responses and is provided with feedback as to the partner's responses, trial by trial, the subject can procure additional information as to his partner's response habits with each succeeding trial. To profit from this available information the subjects would seem to have to attend to and, perhaps, adjust to the partner's behavior. Thus TCGs have been thought to relate to the subjects' abilities to anticipate the viewpoint of the other in social situations. Consistent with the viewpoint that schizophrenia is largely caused by social failures of this type, Solomon (1960) found TCG performance to be worse in schizophrenics than normals, and worse in regressed than in remitted schizophrenics (1966).

Piaget suggests that children between the ages of 9 and 12 have difficulty with tasks

which oblige them to place themselves at points of view other than their own. Such tasks "... are difficult because they necessitate reasoning of a relative and formal kind ... [Piaget, 1959, pp. 64-65]." Flavell (1963, p. 400), in summarizing the results of several studies on the ability of children to discriminate the role attributes of others, states:

The data indicate, first, that young children are, as Piaget has said, very inept in cognitive-interpersonal skills of this genre; and second, that there is considerable growth in such skills during middle childhood and early adolescence.

The present study examined the relative TCG performance of preadolescent and adolescent subjects and compared their performances with previously reported college subjects (Fry, 1965). If Piaget's observations are correct, the preadolescents should find themselves at a considerable disadvantage in the TCG when compared to older groups because of their inability to see the points of view of others. Adolescent subjects should show marked improvement in the game situation since their abilities to see other points of view should be evolving rapidly. College subjects, when compared to younger subjects, should perform at consistently high levels.

Solomon (1966) stressed a particular strategy by which pairs of subjects could quickly solve two-person, three-choice TCGs. If one subject would establish a consistent pattern of choices (cue), the other subject could observe and recognize (track) the

¹ This research was supported by United States Public Health Service Grant MH 07176-01 from the National Institute of Mental Health. The author wishes to express his gratitude to the students and staff at the Belfield School, Charlottesville, Virginia, for their participation in and cooperation with this experiment.

partner's response pattern and modify his own responding to correspond with his partner's pattern. Solomon found that schizophrenics' TCG failures lie in their inability to adopt the tracking rather than the cueing role. Schizophrenics seemed inattentive to their partner's behavior. Using normal college subjects, Fry (1965) predicted and found that partners with discrepant ascendance scores outperformed partners of similar ascendance. He reasoned that establishing the reciprocal tracking and cueing roles would facilitate performance, and suggested that this role relationship could more easily be established between an ascendant and a submissive partner than between either two ascendant or two submissive partners. A similar prediction was made in the present investigation. It was anticipated that among the adolescents and pre-adolescents, partners of discrepant ascendance would perform better than partners of similar ascendance in the TCG.

METHOD

Subjects

Twenty-four fourth graders (9-10-year-olds) and 24 eighth graders (13-14-year-olds) participated and were compared with 36 college subjects. All subjects were paid \$2 for the 2-hour game-play session. IQs were roughly equated at the three age levels (mean IQ = 122). All participants played the game with partners of the same age and sex. Sex was not further controlled, but a subsequent analysis of game performance by sex found no significant differences at any age.

Ascendance

The college subjects had been assigned to quartile levels of ascendance based on the Allport and Allport (1928) A-S Reaction Study. As two grade-school teachers shared the teaching responsibilities for each grade, they served as the raters. Each teacher familiarized herself with the Allport and Allport test and was asked to assign her students to quartile levels of ascendance in accord with the test. Teachers reported 80% or better agreement on ascendance assignments. Where disagreement was found the teachers resolved the disagreement in discussion and then provided the experimenter with the final quartile ascendance assignments. Six grade-school subjects were selected from each quartile list.

Procedure

Subjects met for one 2-hour session in sets of four members, with one set member from each of the four ascendance levels. Each set included only

members of the same age and sex. Game play took place in pairs, and while a given subject played the game with a second set member, the third and fourth set members played the game with each other. Each subject played nine games, three successive games with each of the other three set members. The specific order of subject pairings was determined by ascendance levels in a Latin square design permitting examination of ascendance effects.

Each subject was placed in a booth adjoining his partner's. Each was given an identical set of three different objects which he placed in his own order on the counter in his booth, invisible to his partner. Typical objects might be a door key, a pencil, and a bottle top. The experimenter directed the subjects to pick up one of the three objects so as to match the object selected by his partner. On order from the experimenter both members of the pair extended the selected object forward far enough to permit both members to view both objects in front of the common wall of the adjacent booths. If both objects were the same a *match* occurred, and both subjects were reinforced by a point of score. If a *mismatch* took place both subjects lost a point. After the first match-attempt trial the experimenter introduced a second trial, then a third, etc. Trials continued until Trial 100 or until the pair reached a solution and matched on 10 successive trials. After the first game, second and third games were played with the same partner but with different sets of objects. After a pair reached a game solution, the experimenter passed both pair members notes indicating the game-solution pattern they had used and restricting them from reusing that solution pattern with that particular partner. By the term game-solution pattern here is meant a schematic and repeating sequence of object selections. No pair reached a game solution without using an obvious solution pattern. Although some rather elaborate solutions were found, those most commonly used were (a) selecting one of the objects over and over again, (b) alternately selecting one of two objects, and (c) selecting each object in turn and rotating through the three objects. After the completion of the third and sixth game, the subjects switched partners. Each time they played their first of three games with a new partner the prohibitions against previously used game-pattern solutions were lifted. With each new game, however, new sets of objects were introduced so that subjects had no previous familiarity with the objects in the game situation at the beginning of a game. Subjects competed for a prize which was awarded to the subject with the fewest total mismatches at the end of the experiment.

RESULTS

Figure 1 summarizes the results and shows the mean number of mismatches made in each game by the various age groups. The lack of homogeneity of variance at each age required nonparametric age-group compari-

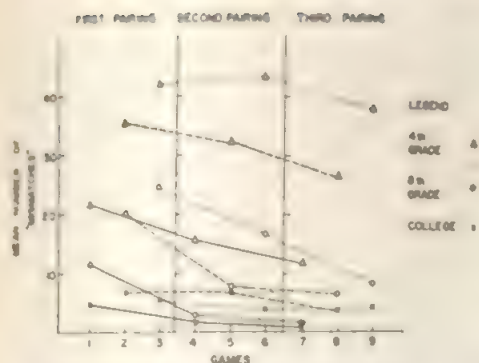


FIG. 1. Mean number of mismatches for each game at each age. (Solid, broken, and dotted lines, respectively, connect points representing the first, second, and third games played with a particular partner.)

sons. Kruskal-Wallis analysis of variance by ranks found all age groups to be significantly different from each other (p 's $< .001$). The college group showed the best overall performance, and the fourth-grade group showed the worst.

Within each age level two analyses of variance were run, one on pair scores and one on individual scores. The analyses of pair scores are presented in Table 1. These $2 \times 3 \times 3$ analyses examined the performance of similar as opposed to discrepant ascendance score pairs (A-S discrepancy effect), as they played their first, second, and third game (problems effect) with each of several partners (pairings effect). The $4 \times 3 \times 3$ analyses of the

individual scores in Table 2 permitted between-subjects examinations of the performance of subjects assigned to the four quartiles of ascendance, and also allowed tests of problems and pairings effect. It should be pointed out that the individual subject scores are not entirely independent of one another, but their analysis does permit the interesting examination of the simple ascendance levels effects.

Pairing Effect

Significant improvement in the performance of college subjects was indicated as they progressed from partner to partner ($p < .06$ in the pair, $p < .005$ in the individual analyses of variance). Substantial improvement was shown by the eighth graders ($p < .001$ on both variance analyses). The fourth graders, however, failed to show any significant improvement as they progressed from partner to partner. Even the apparent improvement on the first problem with each pair failed to reach significance.

Ascendance Effects

The performance of similar versus discrepant ascendance score pairs was examined by comparing the pair performance of subjects playing with partners from adjacent or from more remote quartile levels of ascendance. Table 1 shows that in the college group discrepant ascendance pairs outperformed similar ones ($p < .035$). This was anticipated

TABLE 1
ANALYSES OF VARIANCE OF PAIR MISMATCH SCORES FOR FOURTH GRADE,
EIGHTH GRADE, AND COLLEGE GROUPS

Source	df	4th grade		8th grade		College	
		MS	F	MS	F	MS	F
Pairings (A)	2	543.4	1.20	1726.0 *	9.11***	108.4	2.88
A-S discrepancies (B)	1	51.1	.11	264.5	1.39	172.2	4.57*
Problems (C)	2	3578.4	11.86***	1072.0	5.66**	156.1	4.14*
A \times B	2	638.1	1.39	637.9	3.36*	1.1	.03
A \times C	4	55.0	.12	64.0	.34	19.9	.53
B \times C	2	155.1	.34	64.3	.34	28.5	.76
A \times B \times C	4	1311.9	2.89*	156.5	.84	68.7	1.82
Error	90 ^a	453.4		187.4		37.7	

^a The error term for the college group had 144 degrees of freedom.

* $p < .05$.

** $p < .01$.

*** $p > .001$.

TABLE 2
ANALYSIS OF VARIANCE OF INDIVIDUAL MISMATCH SCORES FOR FOURTH
GRADE, EIGHTH GRADE, AND COLLEGE GROUPS

Source	df	4th grade		8th grade		College	
		MS	F	MS	F	MS	F
Between Ss	23*						
A-S levels (B)	3	21	.01	369	1.56	25.7	.86
Error between	20*	1543		236		29.8	
Within Ss	192*						
Pairings (A)	2	187	.52	3452	15.85**	216.5	5.372*
Problems (C)	2	10731	30.06**	2144	9.84**	312.0	7.74**
A × B	6	443	1.24	235	1.08	24.1	.85
A × C	4	373	1.05	128	.59	40.0	.99
B × C	6	281	.79	161	.74	13.3	.33
A × B × C	12	252	.71	56	.26	24.1	.60
Error within	160*	160		218		40.3	

* College groups' degrees of freedom were between subjects, 35; error between subjects, 32; within subjects, 288; error within subjects, 323.

* $p < .005$.

** $p < .001$.

because of the detrimental effects caused by subjects of similar ascendance attempting to establish a leader-follower relationship. This effect was not found among the grade-school subjects. However, the ascendance pair discrepancy effect was found to interact with the pairings effect in the eighth graders ($p < .05$). This interaction was such that discrepant ascendance pairs outperformed similar ones in the early pairings, but performed somewhat less well in the last pairing. In the fourth grade a significant Ascendance Discrepancy × Problems × Pairings interaction was obtained ($p < .035$).

Table 2 shows the complete failure of the absolute level of ascendance effects to reach significance at any of the ages examined. The absolute level of ascendance variable also failed to interact significantly with other variables.

Problems Effect

The only effect found to be significant at all ages was the problems effect (see Table 1). Unlike the college group, which performed best on the first problem and worst on the second problem, the grade-school children showed a steady decrement in performance from the first to the third problem solved with a partner. In all probability the prohibitions against reusing patterns with a given partner brought about this effect. Simpler

solutions could be expected to be used first, therefore making subsequent solutions more difficult.

Convergence Phenomenon

The college group was observed to develop pattern and response biases among players as to which patterns should be used, in what order they should be used with a partner, and even which object should be selected first in the very first trial of a game (all chi-square p 's $< .05$). There was no evidence to suggest that fourth or eighth graders learned to anticipate their partners' first game choice. The eighth graders seemed to learn to use the same three solution patterns as the college subjects (repeatedly selecting one object, alternating between two objects, and rotating through the three objects), but they did not invoke this sequence of patterns as college subjects had. The fourth graders did not show these convergence phenomena.

DISCUSSION

The results demonstrate the predicted relationships between age and game performance. While college subjects showed moderate improvement as they progressed from partner to partner in the game situation, their performance level was quite high from the beginning. Eighth graders began their game performing at a marked disadvantage when compared to

the college group; however, they showed rapid improvement with practice, thus demonstrating their readiness to improve in the skills demanded. This rapid improvement in game performance is consistent with Piaget's developmental observations outlined in the introduction. Eighth-grade adolescents should be rapidly acquiring skills related to attending to and appreciating the points of view of others, and successful game performance seems to require these skills. Also consistent with Piaget's viewpoint is the fact that the fourth graders (preadolescents) performed very poorly in the game and showed no signs of improving with practice. These failures did not seem due to their inability to invent more complex patterns (e.g., complex patterns such as ABACABAC . . . were occasionally established by fourth graders), rather fourth graders seemed unable to get known patterns established. On 13 occasions fourth-grade pairs failed to reach a game solution when at least one of the partners had previously reached a now legitimate solution with a previous partner. This situation occurred only twice with eighth graders and never among college subjects.

Consistent with the college findings, subjects' absolute level of ascendancy was not found to significantly influence game performance at any age tested. Like the college subjects, discrepant ascendancy score, eighth-grade pairs outperformed similar ascendancy score pairs in the early pairings. Unlike the college subjects, however, this effect appeared to wash out. By the last pairings discrepant

ascendancy score pairs performed somewhat worse than similar pairs. This finding is not without some precedent in the TCG literature. Forsyth, in an unpublished study done in this laboratory, found reciprocal sociometric choice to benefit TCG performance in fourth graders' attempts to play a single game with a first like-aged partner, but not with a second.

The failure of the fourth graders to show the anticipated ascendancy score discrepancy effect is not too surprising in view of the rather general fourth-grade failure in the present TCG situation. Fourth graders failed to find game solutions in 41% of the games they attempted.

REFERENCES

- ALLPORT, W., & ALLPORT, F. H. *A-S reaction study*. Boston: Houghton Mifflin, 1928.
- FLAVELL, J. H. *The developmental psychology of Jean Piaget*. Princeton, N. J.: Van Nostrand, 1963.
- FRY, C. L. Personality and acquisition factors in the development of coordination. *Journal of Personality and Social Psychology*, 1965, 2, 403-407.
- PIAGET, J. *Judgment and reasoning in the child*. Patterson, N. J.: Littlefield, Adams, 1959.
- SCHELLING, T. C. *The strategy of conflict*. Cambridge: Harvard University Press, 1960.
- SOLOMON, L. Schizophrenic communication: Experimental studies in tacit coordination. Paper read at Eastern Psychological Association, New York, April 1960.
- SOLOMON, L. The influence of varying role requirements and premorbid level of adjustment upon role taking behavior in schizophrenics. Paper read at Eastern Psychological Association, New York, April 1966.

(Received April 25, 1966)

PRIOR DYADIC EXPERIENCE AND MONETARY REWARD AS DETERMINANTS OF COOPERATIVE AND COMPETITIVE GAME BEHAVIOR¹

CHARLES G. McCLINTOCK AND STEVEN P. McNEEL

University of California, Santa Barbara

This study investigates the effects of no prior dyadic experience, hostile, and friendly experience, combined factorially with high and low monetary reward on competitive-cooperative behavior in a 2-choice non-zero-sum game. The game matrix is constructed so that a competitive choice unambiguously reflects the operation of a motive to maximize the difference between own and other's score, as opposed to motives to maximize own or joint gain. 15 dyads of Flemish males were assigned randomly to each of the 6 experimental groups within a 3×2 repeated-measures design (100 trials). The results indicate that both prior dyadic experience and reward factors, as well as the number of times the game is actually played, are important determinants of competitive-cooperative behavior.

Considerable research has been conducted in which the Prisoner's Dilemma Game (PDG) has been employed as an experimental paradigm for investigating cooperative and competitive behavior within the dyad (Gallo & McClintock, 1965; Rapoport & Chammah, 1965; Rapoport & Orwant, 1962). Further investigators have suggested that in game-playing behavior (Deutsch, 1958; McClintock & Messick, 1965), and in bargaining behavior (Fouraker & Siegel, 1963) at least three motives may influence the choices made in situations of social interdependence: (a) maximizing own gain, (b) maximizing joint gain, and (c) maximizing the difference between own and other's outcomes.

In order to more clearly delineate the motive of maximizing the difference between own and other's outcomes from own- and joint-gain maximization, the present authors modified the classic PDG into a Maximizing Difference Game (MDG). The payoff matrix of the latter as illustrated in Figure 1 is so constructed that the motive of maximizing difference would prescribe a competitive choice (A_2 or B_2), whereas own- and joint-

gain maximization prescribe cooperative choices (A_1 or B_1). The latter choices are defined as cooperative insofar as they lead to positive mutual reinforcement; A_2 and B_2 choices are defined as competitive insofar as such responses block or frustrate mutual reinforcement.

In previous studies by the authors (McClintock & McNeel, 1966a, 1966b), subjects receiving low rewards were found to compete more in a MDG situation than those receiving high rewards, and subjects permitted to see both their own and the other player's cumulative point totals were more likely to compete than those seeing only their own cumulative scores. The purpose of the present study was to examine the effects of prior dyadic experience upon the subjects' choices in the MDG under two levels of reward.

The variable of prior dyad experience was included because previous research with the PDG has indicated that such experience may affect the cooperative and competitive behavior of dyad members. Harrison and McClintock (1965) found, for instance, that prior mutual task success or failure or the lack of prior experience can influence the level of subjects' competitive behavior. Subjects in the preceding experiment were assigned to one of three experimental groups: (a) prior success, (b) prior failure, and (c) no prior experience. Half of the subjects in the prior-success and prior-failure conditions

¹ This research was supported under Air Force Office of Scientific Research Grant 768-65, and by a Fulbright Research Fellowship awarded the senior author.

The authors would like to acknowledge the help of Josef M. Nuttin, Monika Gerard, and David M. Messick.

played 60 trials of a PDG immediately following their success or failure experience; the second half played 60 trials of the PDG a week after their initial interaction. In the immediate PDG condition, dyads in the prior-success group were significantly more cooperative in their behavior than those in the prior-failure and no-prior-experience conditions. In the 1-week-delay condition, it was found that dyads in both the prior-success and prior-failure conditions were significantly more cooperative in the PDG than those in the no-prior-experience condition, but these two prior-experience conditions did not differ among themselves. These findings would suggest that success and failure has a differential effect upon subjects' behavior when it immediately precedes the playing of a PDG; after a 1-week interval, prior experience, regardless of outcome, tends to produce more cooperative behavior than no prior experience.

The variable of reward level has been examined both within the PDG (Gallo, 1963; Rapoport & Chammah, 1965), and within the MDG (McClintock & McNeel, 1966a, 1966b). Given the MDG, it is possible to assess the behavioral effects of reward level upon the relative dominance of the motive to maximize difference since it can be assumed that competitive responses are made for this purpose. One might expect that low monetary rewards would produce more competitive behavior than high monetary rewards. Under conditions of low reward, the amount of money foregone for competing is small. Conversely, given high rewards, the motives to maximize own and/or joint gain should become more salient and result in more cooperative behavior since the regret to both players of competing (attempting to maximize the difference between own and other's score) can be assumed reasonably to increase as a function of the amount of money foregone.

In the present study, prior dyad experience was manipulated by permitting subjects, other than those in the no-prior-experience group, to play a pregame. This pregame consisted of a series of 10 trials in which each subject was given a beginning score on each trial ranging from +7 to -6 points. In addition, on each trial subjects were displayed an ordered series of five negative and/or posi-

tive points, one of which he could contribute to the other player's cumulative score, for example, +5, +3, 0, -3, or -5. Throughout the 10 trials the experimenter, without the subject's knowledge, controlled the amount of points which each subject received from the other subject. The induction to produce an expectation that the other player was friendly was accomplished by informing the subject that the other dyad member had given him predominantly high positive points; an expectation that the other player was hostile was achieved by informing the subject that the other person had assigned him primarily high negative or low positive points among those available. For the prior-experience conditions, both subjects within a given dyad received either a friendly or hostile induction. Reward level was manipulated by assigning either high or low monetary payoffs to each point in the MDG payoff matrix.

Thus the present study employed a 3×2 design with three conditions of prior experience: no prior experience (N), prior experience-friendly (F), and prior experience-hostile (H); and two conditions of reward level: high reward (HR), and low reward (LR). Given this overall design, the following major hypotheses were advanced:

1. Within an MDG the three prior-experience conditions were expected to be ordered as follows in terms of decreasing competitiveness: no prior experience, prior experience-hostile, and prior experience-friendly. Since this experiment resembles the Harrison and McClintock immediate condition in a number of ways, it was hypothesized that the F condition would behave significantly more cooperatively than the H and N conditions, but the two latter conditions were not expected to differ significantly.

2. Consistent with previous findings reported by the present authors, it was hypothesized that subjects in the LR condition would make more competitive responses than those in the HR condition.

Unfortunately, there was no *a priori* basis (theoretical or empirical) from which to make predictions concerning the ordering of the six experimental groups in terms of the amount of competitive behavior displayed. Such an ordering would require *a priori* esti-

mates of the relative effects of variations in prior experience versus reward level. However, since the hypothesized effects of prior experience and reward level operated in the same direction in two of the six groups, the following predictions could be made: subjects in the no-prior-experience-low-reward group (N-LR) should manifest more competitive behavior than the other five experimental groups; subjects in the friendly-prior-experience-high-reward group (F-HR) should manifest less competitive behavior than the other five experimental groups.

METHOD

Subjects

Subjects were 180 male undergraduate Flemish students enrolled in introductory psychology courses at the University of Louvain, Belgium. Dyad members were relative strangers, and within this restraint dyads were randomly assigned to the six experimental groups. Following the experiment, each subject was asked how well he knew the other member of the dyad. Subjects were considered as strangers if they had spoken once in a while previous to the experiment, but "only on a superficial level."

Apparatus

In the pregame, the subjects were provided with information concerning the number of points they had to begin with on each trial, and a range of points which they could assign the other player. The subjects communicated their decision as to the number of points they wished to afford the other player by pushing one of five buttons labeled from A to E. After each subject had made his choice, the response supposedly made by the other (experimenter actually intervened) was displayed to each subject by the lighting of one of five red lights corresponding to the point options of the other player. Then the number of points which each subject had received was added to two digital scoreboards mounted in front of each subject. For each subject one scoreboard kept track of the subject's own cumulative score; the other, of the other subject's cumulative score.

In the MDG, the experimental apparatus consisted of the following equipment for each subject: (a) a 2×2 matrix (see Figure 1) with lights in each cell to indicate the subjects' joint choice, (b) two push buttons for each subject by which Subject 1 indicated his choice of row in the matrix, and Subject 2 his choice of column, (c) a red start signal indicating the beginning of each trial, (d) two digital display boards which provided the subjects with a continuous cumulative record of their own and the other player's total points, and (e) a card

mounted in front of the subject indicating the monetary reward associated with each point value in the matrix. A copy of the matrix was also posted in such a position that both subjects could view it simultaneously, and thus assure themselves that they were both playing the identical game.

General Procedure

The subjects were seated in front of individual game boards on opposite sides of a table. They were visually isolated, by a barrier, and an auditory discrimination study was conducted prior to the experiment to assure that the recording equipment provided no differential information. The subjects were not permitted to talk out loud to the other player or to themselves. All instructions to the subjects were taped, and were identical except for those additions and deletions required by the two independent variables, prior dyad experience and reward level.

As noted previously, subjects who played the pregame were provided with information concerning the other player's choices designed to produce the expectation that the other player was either friendly or hostile. Subsequent to participation in this pregame, the subjects were requested to complete a 17-item semantic differential scale designed to measure their positive or negative evaluation of the other player; that is, the subjects were requested to "describe the other player as you see him," rating him on a number of polar adjectives such as pleasant-unpleasant, helpful-frustrating, and warm-cold. Subjects in the N condition did not participate in the pregame, and hence did not evaluate the other player prior to playing the MDG.

Following the pregame, all subjects were given instructions concerning the MDG. These instructions were neutral in terms of suggesting cooperative or competitive behavior. The nature of the game was explained using two demonstration trials, and it was emphasized that the outcome for each player was dependent upon the joint choices of both play-

		PLAYER II	
		B ₁	B ₂
PLAYER I	A ₁	6 6	0 5
	A ₂	5 0	0 0

MDG MATRIX

FIG. 1. Matrix for MDG. (Outcome for Player 1 is in the upper left corner of the cell.)

as Subjects in the HR groups were informed, "You can make up to 660 Belgian francs" (\$13.20), whereas those in the LR group were told, "You can make up to 33 Belgian francs" (\$.66).² The subjects read an affidavit, signed by a senior member of the department, which testified in very explicit terms that the subject would be able to keep the money he had accrued by the end of the experiment.

The subjects then played 100 trials of the game. As soon as both subjects had responded on a given trial, their mutual choices were displayed in the appropriate cell of the payoff matrix on each of their respective game boards, and the cumulative scores were updated. A red light then signaled that they should begin the next trial. Upon completion of the experiment, subjects completed a questionnaire which included items to ascertain whether they had prior information concerning the nature of the task and were paid the money they had earned during the game.

In the following results section, we will consider first evidence concerning the effectiveness of the pregame induction, and then present those results which bear upon the hypotheses and predictions of the study.

RESULTS

Pregame Induction

To determine whether the pregame induction had the desired effect of producing a differential image of the characteristics of the simulated other player, a 17-item semantic differential scale was administered to the subjects immediately following the pregame. It was found that subjects interacting with a simulated friendly other evaluated the other significantly more positively on the semantic differential scale than those interacting with a simulated hostile other ($t = 6.40, p < .0005$).

A second method for assessing the success of the induction involved analyzing the total number of positive and negative points which the subject afforded the simulated other player across the 10 trials of the pregame. It was observed that those playing with a simulated friendly other gave the other player a mean of 16.72 out of a possible

point total ranging from -74 to +71 points; those in the hostile induction afforded the other player a mean of -12.29 points ($t = 5.82, p < .0005$). Thus the pregame inductions produced significant differences both in subjects' evaluations of the other player, and in the subjects' behavior toward the other player during the pregame.

In order to determine whether a relationship existed *within* the two simulation conditions between the behavior of subjects (the points given to the other player) and the subjects' evaluations of the other player, appropriate correlation coefficients were computed. It was observed that there was some tendency within the simulated friendly other condition for those who afforded the other player more points during the pregame to evaluate him more positively following the pregame ($r = .23, p < .10$). Within the simulated hostile other condition, those subjects who gave the other player more points during the pregame evaluated him more negatively following the pregame ($r = -.43, p < .001$). The latter would seem to indicate that persons who act more generously, in terms of giving points (rewards) to a hostile other, are subsequently likely to evaluate that person more negatively when their generosity is not reciprocated by a hostile other than are persons who provide less in the way of positive rewards.

Major Analyses

In the present study, competitive rather than cooperative responses are utilized as the basic dependent variable in the major analyses. This is in contrast to previous PDG studies where the subjects' behavior is described in terms of cooperative responses. This change in emphasis is suggested by the nature of the MDG matrix in which a single motive, maximizing the difference between own and other's score, is assumed to underlie competitive choices.

Conditions analysis. The primary analysis performed was a 3×2 repeated-measures analysis of variance. The number of individual competitive responses (A_2 or B_2) occurring within dyads per 10-trial block served as the dependent variable. Table 1

² These totals are based on 110 trials, of which the first 100 are employed in the present analysis. The "you" in the instruction was the impersonal form, and subjects' potential winnings were made up of the points which each could win, summed for the HR group, and taken individually for the LR group. This technique was employed to maximize the perceived reward differential at minimum cost to the experimenter.

TABLE 1

ANALYSIS OF VARIANCE OF COMPETITIVE RESPONSE
FOR 100 TRIALS: PRIOR EXPERIENCE \times REWARD
LEVEL \times TRIALS

Source	df	Adjusted df	MS	F
Between	89			
Prior experience (P)	2		638.78	4.06**
F vs. H and N	1		840.50	5.34**
H vs. N	1		433.50	2.75*
Reward (R)	1		524.41	3.33*
P \times R	2		39.52	.25
Error (between)	84		157.53	
Within	810			
Trials (T)	9	1	64.78	5.24**
P \times T	18	2	29.56	2.39*
R \times T	9	1	37.63	3.05*
P \times R \times T	18	2	14.52	1.17
Error (within)	756	84	12.36	

Note.—Results from Box's multivariate analogue of Bartlett's test for homogeneity of variance rule against the pooling of covariances (see Winer, 1962, pp. 369-374). As a result, a conservative procedure (suggested by Greenhouse & Geisser, 1959) which does not require this assumption was used to obtain significance values. This procedure involves appropriate adjustment of the degrees of freedom associated with each MS in the within section of the analysis.

* $p < .10$.

** $p < .025$.

presents analysis of variance results for the total 100 trials. In addition, separate analyses of variance were run on the first 30 and the last 70 trials in order to more clearly delineate changes in strategy through time.³ The mean proportions of competitive responses by prior experience and reward conditions are presented in Table 2. As hypothesized, there was a significant overall effect due to prior dyadic experience. Appropriate orthogonal comparisons indicate that this effect was primarily due to a significant tendency for the F condition to be more cooperative than the H and N conditions. There was, however, only a marginal tendency for the H condition to be more cooperative than the N condition.⁴ Thus, the data verified Hy-

³ The data were divided, a priori, in this manner on the assumption that subjects spend the first few trials of the game "learning" how to play within their particular experimental manipulation. This type of subanalysis was performed because extreme heterogeneity legislated against the appropriateness of using a pooled estimate of within-group error to make comparisons between conditions or groups. It was necessary to use the conservative procedure (see note to Table 1) for estimating significance levels in the 70-trials analysis, but usual procedures were applicable to the 30-trials analysis.

⁴ A second orthogonal comparison indicates that prior experience per se (whether hostile or friendly)

TABLE 2

MEAN PROPORTIONS OF COMPETITIVE RESPONSE
PRIOR EXPERIENCE, REWARD LEVEL, PRIOR
EXPERIENCE \times REWARD LEVEL

Source	100 trials	1st 30 trials	Last 70 trials
Grand mean	.625	.590	.640
Prior experience (P)			
No experience (N)	.700	.605	.745
Hostile (H)	.615	.600	.620
Friendly (F)	.555	.570	.550
Reward (R)			
Low (LR)	.660	.595	.690
High (HR)	.585	.585	.585
P \times R			
N-LR	.760	.610	.825
N-HR	.645	.600	.660
H-LR	.640	.625	.645
H-HR	.590	.575	.595
F-LR	.585	.585	.600
F-HR	.525	.555	.500

pothesis 1. The overall reward effect, however, was only marginal; the LR condition displayed a somewhat higher propensity to compete than the HR condition. Both prior experience and reward effects were strong in the last 70 trials (prior experience: $F_{2,70} = 5.31$, $p < .01$; reward level: $F_{1,70} = 4.63$, $p < .05$), but negligible during the first 30 trials.

The trials effects, which reflect changes in strategy through time, are particularly interesting. There was a significant overall tendency for dyads to behave in a more competitive fashion as the game progressed. The 30- and 70-trial subanalyses indicate that this trials effect took place primarily in the first 30 trials ($F_{9,30} = 27.99$, $p < .001$; $F_{70} < 1$).

Using a very conservative estimate of significance (see Note, Table 1), a marginal overall interaction ($p < .10$) was found between prior experience and trials, even though this interaction was nonsignificant in both the 30- and 70-trial subanalyses. It would appear that the different prior-experience conditions generated different response patterns during the game. Figure 2A indicates that after an initial orientation to the game (first trial block), the behavior of the three

was the significant factor ($F = 5.68$, $p < .025$) as opposed to the type of prior experience ($F = .46$, ns).

prior-experience conditions was related in the same way as the groups in the Harrison and McClintock immediate condition (Trial Block 2), the F condition being much more cooperative than either the H or N conditions, which did not differ. Then, beginning with Trial Block 3, the three prior-experience conditions exhibited behavior much more similar to the Harrison and McClintock delay condition. Thus, after playing 20 trials of the game, the N condition consistently showed the greatest tendency to compete, the F condition was consistently the least competitive, and the H condition fell between these two extremes.

Using the conservative method of estimating significance noted previously, the Reward \times Trials interaction was also marginally significant ($p < .10$), and it is evident that this effect also derived primarily from the first 30 trials ($F_{30} = 3.04$, $p < .05$) rather than the last 70 trials ($F_{70} = 1.97$, ns). The same general trend occurred here as in the prior-experience conditions (see Figure 2B); there was an initial increase in competitiveness followed by a divergence of the conditions, the LR condition becoming slightly more competitive and the HR condition increasingly less competitive as the game

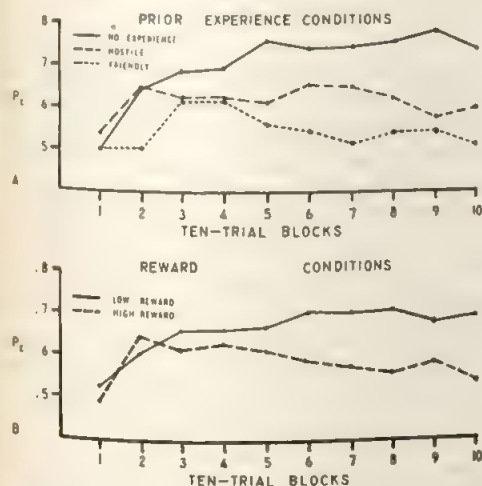


FIG. 2. Proportion of competitive responses by 10-trial blocks for the prior-experience (A) and reward (B) conditions. ($N = 30$ dyads for prior-experience conditions, and $N = 45$ for reward conditions.)

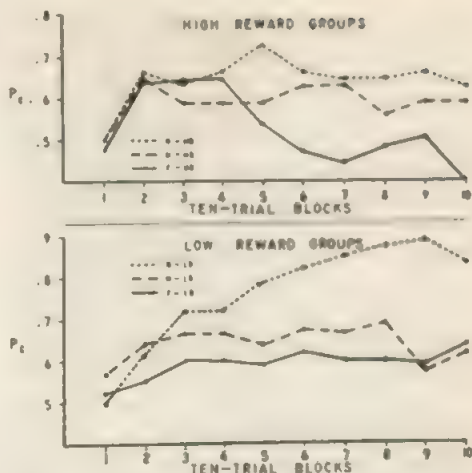


FIG. 3. Proportion of competitive responses by 10-trial blocks for the individual experimental groups. ($N = 15$ dyads per group. Legend: N = no prior experience, H = hostile prior experience, F = friendly prior experience, LR = low reward, and HR = high reward.)

continued. In addition, as in the prior-experience conditions, the predicted ordering of the reward conditions (LR more competitive than HR) occurred without exception after the first two trial blocks. It is evident, then, that both prior experience and reward factors affected differential changes in dyads' competitive behavior over time.

Groups analysis. The interaction between prior experience and reward level was insignificant in all major analyses. This indicates that the behavior of each individual group could be accounted for reasonably by the main effects of prior experience and reward. This result can be seen clearly in the ordering of the mean proportion of competitive choices for the six experimental groups (see Table 2).

Figure 3 presents the trends in choice behavior for each of the six experimental groups as the game progressed. The graphs indicate no differential trials effects among groups except for the N-LR and F-HR groups (the two groups in which both manipulations would be expected to influence behavior in the same direction). The interaction of groups with trials did not reach significance in any of the major analyses, presumably due to the marked similarity of the profiles of the other

four groups. It seems evident, however, in agreement with the predictions, that the behavior of the N-LR and F-HR groups did differ from the other four groups and from each other as a function of time. Namely, after an initial increment in competitive behavior in all groups, four of the groups remained relatively constant in terms of competitive behavior, whereas the N-LR group showed a substantial increment in competitive behavior over trials, and the F-HR group showed a substantial decline in competitiveness.

Supplementary Analyses

Variance. Since the dependent variable for the major analyses was a dyad score, there is a great deal of information regarding subject variability which is not considered by this type of analysis, that is, information dealing with the variation in behavior occurring within dyads. Messick and McClintock⁵ have discussed a method for investigating intradyad as well as interdyad variation. This method involves partitioning the total variation occurring in any given group into two independent additive categories, within-dyad variation (variation of subjects in a dyad about the dyad mean) and between-dyad variation (variation of dyad scores about the mean of the group or condition). When divided by the appropriate degrees of freedom, each of these measures yields a variance measure.⁶

Variance between dyads. The variance between dyads (VBD) in a given condition (or

⁵ D. M. Messick and C. G. McClintock, "Measures of Homogeneity in Two-Person Two-Choice Games," unpublished manuscript, 1966.

⁶ This partition of variation, basic to the analysis of variance, is presented in Winer (1962, p. 51). In the present context, the partition of the total variation occurring in a group of dyads is as follows:

$$\sum_{j=1}^n \sum_{i=1}^2 (X_{ij} - X_{..})^2 \\ = \sum_{j=1}^n \sum_{i=1}^2 (X_{ij} - X_{.j})^2 + 2 \sum_{j=1}^n (X_{.j} - X_{..})^2,$$

where X_{ij} is the number of competitive choices of individual i in dyad j , $X_{..}$ is the group mean number of competitive choices, and $X_{.j}$ is the mean number of competitive choices of dyad j .

group) is a measure of the variability of the dyad scores with respect to the mean of that condition (or group). As a rough test of time effects in VBD, the VBDs for the first and last trial blocks of each condition were compared statistically.⁷ The difference between the VBD on these two trial blocks was highly significant for the HR condition ($t = 4.18$, $p < .001$), but negligible for the LR condition. In particular, it appears that dyads in the HR condition were significantly more heterogeneous in behavior at the end of the game than at the beginning, whereas dyads in the LR condition remained relatively homogeneous in behavior throughout. In the prior-experience manipulation, the difference between the VBD on Trial Blocks 1 and 10 was significant only within the H condition ($t = 3.16$, $p < .01$). The N condition, however, did reach marginal significance ($t = 1.92$, $p < .10$), whereas the difference between these variances for the F condition was nonsignificant. Thus, comparing dyads in the different prior-experience manipulations, only the H condition became significantly more heterogeneous in behavior as the game continued, though a similar tendency was observed in the N condition. It is evident, then, that the major manipulations of prior experience and reward are differentially effective in controlling interdyad variability.

Figure 4 presents VBD (top portion of each graph) for each of the six experimental groups plotted as a function of 10-trial blocks. The differential time effect which developed across groups is evident. As in the conditions analysis, the VBDs of each group for Trial Blocks 1 and 10 were compared. As we would expect from the main effects of the previous paragraph, the difference between the VBD on these two trial blocks was significant only for the H-HR group ($t = 3.84$, $p < .01$), although each of the other HR groups did reach marginal signifi-

⁷ The analyses to follow would be a priori in nature and certainly less subject to Type I error if there were an appropriate omnibus test (with a repeated-measures factor) applicable to variances. Such a test does exist for a single factor design without repeated measures (see Scheffé, 1959, Section 3.8), but we know of no extension of this test to the repeated-measures case.

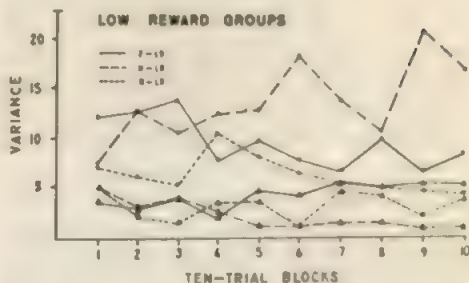
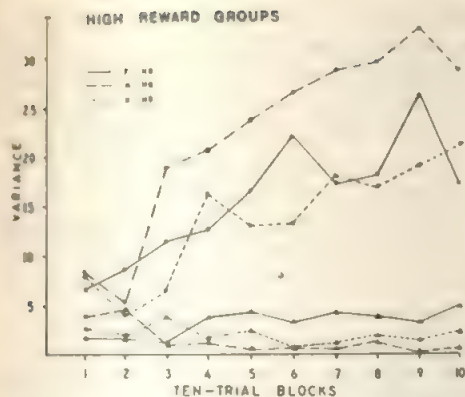


FIG. 4. Variance between dyads (data points are dots) and variance within dyads (data points are triangles) by 10-trial blocks for the individual experimental groups. ($N = 15$ dyads per group. Legend: N = no prior experience, H = hostile prior experience, F = friendly prior experience, LR = low reward, and HR = high reward.)

cance (N - HR , $t = 1.85$, $p < .10$; F - HR , $t = 1.79$, $p < .10$). Thus, of the six experimental groups, only the H - HR group was significantly more heterogeneous in behavior at the end of the game than at the beginning. Furthermore, it appears that a major portion of the significant increase in between-dyad heterogeneity in the HR and H conditions was due to the increase in the H - HR group.

In conjunction with the above dyadic variance effects, it is instructive to note again the ordering of the individual experimental groups in terms of competitive behavior (see Table 2). Comparing the six groups, it is evident that in each major prior-experience category, the LR group showed the least variability between dyads and the most competitiveness. Conversely, the HR group in each prior-experience category showed the most variability and the least competitiveness. Clearly, then, this inverse relationship between variability and competitiveness was due to the major reward manipulations, regardless of the prior-experience variable.⁸

Variance within dyads. Though the variance within dyads (VWD) is not accounted for in the analysis of variance, it is perhaps the most interesting form of variability from

a social psychological viewpoint. This measure considers the responses of individuals rather than averaged responses of two dyad members, and it reflects differences between the behavior of dyad members in the group. Figure 4 presents VWD for each group (bottom portion of each graph) plotted as a function of 10-trial blocks.

Two points are of particular interest in the VWD data: (a) the markedly low scores which characterized all conditions and groups, and (b) time effects which occurred in spite of these low scores. Time effects in VWD were tested in the same manner as VBD. In the two reward conditions there were no significant differences between the VWD scores for the first and the last trial blocks. Considering the prior-experience conditions, the only significant difference was a strong trend for those in the H condition to become more homogeneous in their behavior within dyads ($t = 4.46$, $p < .001$).

As would be expected from the above main effect, the two H groups, considered separately, decreased in VWD from the first to the tenth trial block. This decrease, however, was significant only in the H - LR group ($t = 2.94$, $p < .02$). On the other hand, both of the F groups increased in VWD, and this increase was marginally significant in the F - HR group ($t = 2.12$, $p < .06$). No significant effects occurred in the N groups. Thus, in general, the H groups attained and then

⁸ It should be noted that the increases in VBD cannot be attributed to between-group differences in proportion of competitive responses, since the expected variances for the obtained 10-trial block proportions are small and of the same general magnitude.

maintained extremely homogeneous behavior within dyads, whereas the F groups were considerably more heterogeneous, and the N groups fell between these two extremes. However, the most striking fact about the VWD data is the clear indication that very little variability occurred in individuals' choice behavior *within* dyads.

Comparison of VBD and VWD. Since VBD is a measure of interdyad similarity, it is clear that dyads in the H-LR groups as well as dyads in all HR groups decreased in similarity as the game continued. Dyads within these experimental manipulations apparently settled on progressively different solutions to the game, whereas the dyadic solutions to the game in the N-LR and F-LR groups became progressively more similar. On the other hand, VWD, which measures the similarity of members within a dyad, was remarkably low in all groups. The low VWD indicates that, though dyads within a given group may have been settling on different solutions, the rate of change for each subject was very similar to that of his dyad partner (i.e., though the subjects' proportions were changing, the proportions of any given pair remained similar as the change occurred). This means that a high degree of reciprocal action was taking place between dyad members; for example, if a player made one kind of response, the other player was very likely to reciprocate that response on the next trial. Note that this could occur either when the subjects were "locked in" on the same response, or when the subjects were consistently trying to get the better of the other player, and were therefore unwilling to stay on the

"short end" of a unilateral response. Thus, this measure has *at least* two quite different psychological determinants. The subsequent matched-response analysis provides information concerning the relative frequencies of "locked in" responses within dyads across trial blocks (see Tables 3 and 4). A review of such lock-in data (as well as measures of a player's willingness to be on the losing side of a unilateral response, e.g., Rapoport & Chammah's, 1965, "martyr ratio") indicates that the VWD measure is a more complexly determined measure than it might initially appear to be.

Transition probabilities. Table 3 presents the probabilities of transition from one dyadic state to the same or another state for each of the experimental conditions and groups over 100 trials.⁹ These data are presented in tabular form, as the interdependencies between the dyadic transitions make statistical evaluation difficult.

In terms of prior experience, some transition probabilities are ordered as might be expected, given our original hypothesis. For example, the transition to a mutually competitive (A_2B_2) state from an A_2B_2 state or an unstable (A_1B_2 or A_2B_1) state is most probable in the N condition, less probable in the H condition, and least probable in the F condition. Similarly, a transition from a mutually competitive state to an unstable state is least probable in the N condition and most probable in the F condition. Hence, in these instances, dyads in the N condition

⁹ The complete transition matrices for all groups are available from the authors.

TABLE 3
PROBABILITIES OF TRANSITION BETWEEN STABLE AND UNSTABLE DYADIC STATES

Transition	Time		Experimental conditions					Experimental groups					
	t	$t + 1$	N	H	F	HR	LR	N-LR	N-HR	H-LR	H-HR	F-LR	F-HR
Stable-Stable	A_1B_1	A_1B_1	.469	.750	.623	.738	.463	.239	.585	.655	.815	.390	.770
	A_2B_2	A_2B_2	.710	.663	.569	.654	.661	.748	.661	.621	.706	.563	.576
Unstable-Stable	A_1B_2/A_2B_1	A_1B_1	.136	.114	.146	.115	.159	.128	.143	.116	.112	.188	.094
	A_1B_2/A_2B_1	A_2B_2	.455	.450	.319	.412	.412	.447	.462	.466	.431	.286	.352
Stable-Unstable	A_1B_1	A_1B_2/A_2B_1	.368	.205	.310	.221	.389	.410	.347	.269	.160	.499	.191
	A_2B_2	A_1B_2/A_2B_1	.228	.294	.363	.288	.282	.191	.276	.329	.258	.382	.344

TABLE 4

PROPORTIONAL DISTRIBUTION OF MUTUAL GAME CHOICES OVER 100 TRIALS
FOR EXPERIMENTAL CONDITIONS AND GROUPS

Mutual game choice	Experimental conditions					Experimental groups					
	N	H	F	LR	HR	N-LR	N-HR	H-LR	H-HR	F-LR	F-HR
stable (A_1B_1)	.146	.223	.231	.153	.247	.098	.193	.182	.263	.178	.284
stable (A_2B_2)	.548	.452	.345	.478	.418	.618	.477	.463	.441	.354	.337
Unstable (A_1B_2/A_2B_1)	.306	.325	.424	.347	.335	.284	.329	.356	.296	.469	.380

behaved most competitively and those in the F condition least competitively.

However, the data indicate that an important exception to the original hypothesis occurred in the transitions to and from cooperative states in the H condition. Dyads in the H condition were considerably more likely to stay in the mutually cooperative state than those in either the F or the N conditions. A congruent trend can be noted in the transition from an A_1B_1 to the unstable states: the H condition was least likely to make this transition. However, this is to be expected from purely arithmetic considerations, since the probabilities of the two transitions dealt with above, plus the transition from A_1B_1 to A_2B_2 (not reported) must sum to 1. Thus, though dyads in the H condition made fewer mutually cooperative responses than dyads in the F condition over the total game (see Table 4), they were more likely to stay in an A_1B_1 state once such a state occurred than were dyads in the F condition.

In considering the reward conditions, one finds the transition probabilities ordered as would be expected only for the transitions to and from the mutually cooperative A_1B_1 state. In particular, the probability of a transition to the A_1B_1 state was greatest for the HR condition, whereas the probability of moving away from an A_1B_1 state was smallest for this condition. It appears that reward did not affect transitions to and from the A_2B_2 state. In other words, once dyads landed in a mutually competitive situation, the reward variable did not differentially affect the probability of the occurrence of any one of the four possible dyadic states on the next trial. This particular lack of effect suggests that,

given the structure of the present game and the reward manipulations, competition is the more inflexible and stable game strategy.

When transition probabilities are investigated on a group basis, some interesting effects not indicated by the main manipulations are revealed. In general, the prediction that the N-LR and the F-HR groups should be the most and least competitive, respectively, of the six experimental groups was not systematically supported by the transition analysis. The N-LR group was the most competitive group in only three of the six reported transitions. Furthermore, in no case was the F-HR group the most cooperatively predisposed. The failure of these predictions was due in part to the unexpected main effect of cooperation in the H condition, but this does not explain all of the reversals seen in individual group transition probabilities. It seems evident that there is an interaction (as yet undetermined) occurring between the experimental manipulations of prior experience and reward.

Matched responses. Table 4 presents the proportion within each condition or group of each of the four possible matched responses over the total 100 trials. The striking feature of these data is that within all conditions and groups, except the F condition (and its subgroups, F-HR and F-LR), more mutually competitive responses were made than any other joint response, and this by a substantial margin. Further, in each of the exceptions, the unstable matched responses (A_1B_2 or A_2B_1) occurred more frequently. Thus, the matched-response data confirm the notion stated earlier regarding the relative dominance of competitive responses in the present game.

Mutually competitive (A_2B_2) and mutually cooperative (A_1B_1) responses were also analyzed by prior experience and reward level for Trial Blocks 1, 5, and 10 in order to determine whether the relative frequency of these choices varied through time. In the reward conditions there was, with one exception, a consistent change (an increase in A_2B_2 and decrease in A_1B_1 responses) in the pattern of responding over trials. In addition, as predicted, the LR condition made consistently more A_2B_2 responses and consistently fewer A_1B_1 responses than the HR condition. In the three prior-experience conditions, the pattern of responding over trials was not consistent. There was a large increase in A_2B_2 responses and a slight decrease in A_1B_1 responses from Trial Block 1 to Trial Block 5, followed by a slight decrease in A_2B_2 responses and a slight increase in A_1B_1 responses on Trial Block 10.

The probability of a matched response on Trial Block 1 was approximately the same regardless of prior-experience condition or type of matched response. Thus, the ordering hypothesis regarding the three prior-experience conditions (the N condition being the most competitive, and the F condition being the least competitive) was upheld only for

Trial Blocks 5 and 10 with regard to A_2B_2 responses, and only for Trial Block 10 with regard to A_1B_1 responses. Hence, it appears that the effects of a particular prior-experience manipulation did not take effect until the subjects had some experience with the game.

Figure 5 presents the matched-response analysis for the six experimental groups on Trial Blocks 1, 5, and 10. The hypothesis that the N-LR group should be the most competitive and the F-HR group the least competitive of the six experimental groups (on any given trial block) was not upheld unequivocally for either A_1B_1 or A_2B_2 responses. In terms of mutually competitive behavior, the data are consistent with this hypothesis except in the first trial block: for mutual cooperative behavior, the hypothesis was upheld only in the last trial block.

DISCUSSION

It is evident from the results that both reward level and prior experience affect the overall proportion of competitive and cooperative responses in an MDG. Furthermore, one can observe across conditions that there is a significant trend toward more competitive responses—particularly in the first 30 trials. One can note also that both the prior-experience and the reward conditions seem to interact with time (trials) insofar as the N condition becomes more competitive through time than the H or F conditions, and the LR condition becomes more competitive through time than the HR condition. Finally, although there is no significant interaction between reward and prior experience, the two groups in which the manipulations operated in the same direction show the most competitive (N-LR) and the most cooperative (F-HR) behavior, respectively.

In examining the variance between and within dyads, we note a number of findings consonant with our general hypotheses. We observe that during the course of the game the VBD increased significantly for the HR group but not for the LR group. If one assumes that the more likely strategy in the game is competition, and that high reward favors cooperative choices, then one might expect that dyads in the HR condition are

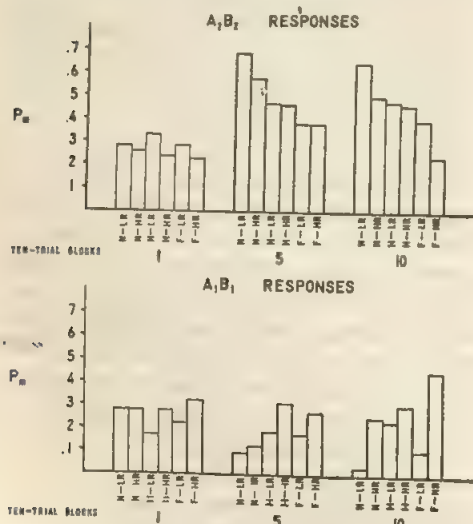


FIG. 5. Distribution of matched responses over time by individual experimental groups. (P_m = proportion of matched responses, $N = 15$ dyads per group.)

in more of a conflict situation than those in the LR condition, and hence dyads in the former condition are more likely to play in more varied ways than those in the latter.

In making VBD comparisons between the prior-experience conditions, one finds that the hostile prior-experience group was the only one to show a significant increase in between-dyad variability across the game. This would seem to contradict the above interpretation, since a hostile induction should be consonant with the assumed tendency to compete in the game, and hence variability between dyads in this condition should decrease or at least remain stable through time. Inspection of each of the six experimental groups clarifies this apparent contradiction, for it can be noted that it is only the H-HR group which shows a significant increase in VBD through time, and it is this group which accounts for the significant variance increase in the H condition. Returning to our earlier interpretation, we can see why dyads in this group should experience the most conflict and hence show the greatest variability in behavior. Namely, the game tends to produce competitive behavior, the hostile induction does likewise, *but* the HR condition makes the cost of competitive behavior high. That the HR manipulation is responsible for creating conflict and variability in dyad behavior can further be seen insofar as the other two high-reward groups, N-HR and F-HR, show an increment in variability between groups across trials which approaches significance ($p < .10$). Conversely, dyads in the LR condition are more competitive, show smaller overall VBD, and manifest no marked increase in VBD across trials.

Examining the VWD, one can note some findings which are consistent with our expectations, and others which are enigmatic. From a common-sense viewpoint, the low VWD in all groups is to be expected, since the dyad members are interacting with each other. However, even with this small variability in choice behavior within dyads, subjects in the F groups were consistently more heterogeneous in their choices within dyads than those in the H groups. It is not immediately apparent why the hostile induction should produce more intradyad similarity in choice behavior

than the friendly induction. One might speculate that the hostile induction and the competitive nature of the game operated together to produce more similar subject expectations within the H-LR and H-HR dyads. However, we are faced with our earlier finding that the group with the highest VBD was the H-HR group. In other words, the H-HR dyads displayed the most heterogeneous patterns of choice behavior between dyads, but the most homogeneous pattern within dyads. It is the latter which is difficult to rationalize.

The small variance in choice behavior within dyads in the H condition leads to an unanticipated finding in the analysis of transition probabilities. Namely, once subjects in the H condition made an A_1B_1 choice, they were considerably more likely to remain in that mutually cooperative state than those in the F condition. This may be partially explained by noting the unexpectedly low likelihood of transition from mutual cooperative to mutual cooperative states in the F-LR group. However, the reason for this finding is unclear, and, as above, the finding is enigmatic given the conceptual orientation of the present study, and the general consistency of other findings with this orientation.

In considering the joint choices of subjects on a given trial, we find that there is a higher likelihood across conditions and groups for an A_2B_2 choice than for any of the other three possibilities. In examining the reward conditions, we observe that there are more mutually competitive responses and fewer mutually cooperative responses in the LR than the HR condition. The prior-experience conditions produced results consistent with expectations except in one instance: as expected, the N condition displayed the highest proportion of mutually competitive responses on Blocks 5 and 10, and the smallest proportion of mutually cooperative responses on Trial Blocks 5 and 10; the F condition produced the lowest proportion of A_2B_2 responses on Trial Blocks 5 and 10, but there was no difference between the H and F groups in terms of mutually cooperative (A_1B_1) responses. This latter exception is of course not independent from the earlier unexpected findings, that is, the H condition showing the lowest VWD, and being the condition

with the highest likelihood for remaining in a mutually cooperative (A_1B_1) state.

In general, the overall findings of the present study are consistent with previous findings reported by the present investigators. Namely, a sizable proportion of subjects' responses in the MDG reflect an orientation toward maximizing the difference between own and other's outcome even when another choice dominates in terms of own and joint maximization of outcome. Furthermore, we have again demonstrated that the degree to which subjects play to maximize the difference between own and other's outcomes (defined as a competitive response in the present game) can be manipulated. In the present study, both prior experience and reward level affected the behavior of the subjects as reflected in the proportion of competitive choices made by dyads, the variance between and within dyads, the probabilities of moving from one dyadic state to another, and finally in the proportion of matched choices of particular types.

Finally, an important problem which the present study points out is that, while game situations provide a useful experimental paradigm for investigating certain kinds of social interaction, the analytic techniques available for evaluating such data are incomplete and insufficient. More complete analytic methods are needed in order to take into account all of the available data, therefore avoiding Type I errors, and enabling the researcher to talk about both social and statistical interactions in a meaningful way.

REFERENCES

- DEUTSCH, M. Trust and suspicion. *Journal of Conflict Resolution*, 1958, 2, 265-279.
- FOURAKER, L. E., & SIROEL, S. *Bargaining behavior*. New York: McGraw-Hill, 1963.
- GALLO, P. S. The effects of different motivational orientation in a mixed motive game. Unpublished doctoral dissertation, University of California, Los Angeles, 1963.
- GALLO, P. S., & MCCLINTOCK, C. G. Cooperative and competitive behavior in mixed-motive game. *Journal of Conflict Resolution*, 1965, 9, 68-78.
- GREENHOUSE, S. W., & GEISSER, S. On methods in the analysis of profile data. *Psychometrika*, 1960, 24, 95-112.
- HARRISON, A. H., & MCCLINTOCK, C. G. Previous experience within the dyad and cooperative game behavior. *Journal of Personality and Social Psychology*, 1965, 1, 671-675.
- MCCLINTOCK, C. G., & MCNEEL, S. P. Reward and score feedback as determinants of cooperative and competitive game behavior. *Journal of Personality and Social Psychology*, 1966, 4, 606-613 (a).
- MCCLINTOCK, C. G., & MCNEEL, S. P. Reward level and game playing behavior. *Journal of Conflict Resolution*, 1966, 10, 98-102. (b)
- MCCLINTOCK, C. G., & MESSICK, D. Non-zero-sum games—a conceptual orientation. In C. McClintock (Chm.), *Game theory and social behavior: Symposium presented at Western Psychological Association*, Honolulu, April 1965.
- RAPAPORT, A., & CHAMMAH, A. *Prisoners' dilemma: A study of conflict and cooperation*. Ann Arbor: University of Michigan Press, 1965.
- RAPAPORT, A., & ORWANT, C. Experimental games: A review. *Behavioral Science*, 1962, 7, 1-37.
- SCHEFFÉ, H. *The analysis of variance*. New York: Wiley, 1959.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received April 27, 1966)

AN EXPERIMENTAL STUDY OF THE REDUCTION OF HOSTILITY THROUGH FANTASY¹

ANN ROTH PYTKOWICZ, NATHANIEL N. WAGNER, AND IRWIN G. SARASON

University of Washington

This research concerned the relative effectiveness of 2 types of fantasy activity—daydreaming and writing TAT stories—in reducing experimentally aroused hostility, and the relationship of 2 variables—sex and degree of frequency of daydreaming, to the reduction of hostility through fantasy. Sex was a significant variable in that the experimental conditions did not produce greater catharsis in females. Both daydreaming and TAT fantasy were effective cathartic experiences for male Ss, and the cathartic effects were stronger for males describing themselves as frequent daydreamers. Catharsis through fantasy clearly occurs when the type of aggression measured is direct verbal expression of hostility towards the examiner; when self-ratings of hostility are measured, however, fantasy results in increased hostility scores. This study supports, but modifies, the catharsis theory of fantasy.

The cathartic or drive-reducing function of fantasy has been postulated by Freud (1920, 1925) and others (Dollard, Doob, Miller, Mowrer, & Sears, 1939; Dollard & Miller, 1950) and has been put to empirical test by a number of investigators (Berkowitz, 1960; Estess, 1960; Feshbach, 1955, 1956, 1961; Gordon & Cohn, 1963; Kenny, 1952; Mussen & Rutherford, 1961; Saxton, 1962; Siegel, 1956; Singer & Rowe, 1962). Proponents of catharsis theory have attempted to demonstrate that aggression can be displaced or reduced through fantasy, thereby reducing the strength of the drive or the instigation toward aggressive action. Several experiments have demonstrated that under certain conditions fantasy can be drive reducing when the drive measured is aggression (Estess, 1960; Feshbach, 1955, 1961; Singer & Rowe, 1962).

The discrepant findings in the literature, however, suggest that the process of catharsis is by no means a simple one. The necessary conditions appear to be (a) fantasy content

that is relevant to the nature of the drive aroused—for example, neutral movies do not have a cathartic effect on aggressive drive while hostile movies do (Estess, 1960; Feshbach, 1961)—and (b) arousal of the pertinent drive at the time the fantasy is experienced. Aggressive fantasy presented with no or only minimal prior arousal of aggressive drive sometimes results in subsequent drive arousal rather than drive reduction (Feshbach, 1956, 1961; Gordon & Cohn, 1963; Kenny, 1952; Mussen & Rutherford, 1961; Saxton, 1962; Siegel, 1956).

The type of fantasy experience afforded subjects would also seem to be an important variable. Most catharsis studies have utilized a fantasy experience in which the stimuli were external to the subject, such as movies (Estess, 1960; Feshbach, 1961; Mussen & Rutherford, 1961; Siegel, 1956), writing TAT stories (Feshbach, 1955), listening to stories (Gordon & Cohn, 1963; Kenny, 1952), or dramatic play (Feshbach, 1956). Only one study (Singer & Rowe, 1962) investigated spontaneous fantasy, and in it the design prevented a comparison of the relative effectiveness of spontaneous and elicited fantasy. In the present study it was possible to compare two kinds of fantasy with each other and with a control activity.

There has been little investigation of organismic variables in relation to the cathartic effects of fantasy. One might ask if

¹This article is based upon a dissertation submitted to the University of Washington by the senior author in partial fulfillment of the requirements for the degree of Doctor of Philosophy. It was conducted while the senior author held a clinical training grant from the National Institute of Mental Health. The study received partial support from a grant to the second author from the State of Washington Initiative 171 Funds for Research in Biology and Medicine.

fantasy would not provide a more cathartic experience for some subjects than for others. Recent work has suggested that the ability to engage in fantasy may be a learned skill, which, like language development, is related to both constitutional and environmental conditions and constitutes a dimension of individual difference (Singer, 1966; Singer & McCraven, 1961, 1962). A number of studies have demonstrated a relationship between fantasy activity (usually inferred from human movement responses on the Rorschach test) and various measures of motor control, inhibitory capacity, and problem solving (Singer, 1955, 1961; Singer & Opler, 1956; Singer, Wilensky, & McCraven, 1956). Singer (1961) has also demonstrated a positive relationship between a quantitative measure of fantasy activity and delaying capacity in children. Using a self-report questionnaire to assess daydreaming activity, Singer (Singer & McCraven, 1961; Singer & Schonbar, 1961) concluded that a high frequency of daydreaming is related to a dimension of self-awareness and introspection which permits greater access to the many facets of one's personality. It would therefore seem possible that subjects reporting a high frequency of spontaneous fantasy in their daily lives might have incorporated fantasy activity as an adaptive mechanism for dealing with stress and therefore might utilize a fantasy experience more effectively for catharsis. Thus, in an experimental situation designed to test the drive-reducing function of fantasy, it was expected that the cathartic effects of fantasy would be more pronounced for subjects reporting frequent spontaneous daydreaming than for those who seldom daydreamed.

Another organismic variable which has not been systematically investigated in terms of the cathartic function of fantasy is sex. Most catharsis studies with adult subjects have either exclusively used males or have failed to analyze the data for possible sex differences.

In the present study, male and female subjects, differing in reported frequency of daydreaming, were insulted to arouse aggressive drive. They then spent a period of time either daydreaming, writing TAT stories, or working on a control activity. The subjects were then given an attitude questionnaire, a self-

report hostility scale, and a word-association test to assess their residual level of hostility. Control subjects were given identical treatments but were not insulted. To test for the effectiveness of fantasy in reducing aggressive drive, the experimental subjects' scores on the three dependent variables were compared to the scores of subjects similarly insulted but who had not engaged in a fantasy activity.

It was predicted that subjects provided with an opportunity to engage in fantasy following frustration would show less aggressive feeling directed toward the frustrating experimenter than frustrated subjects having no cathartic experience. It was also predicted that subjects who reported themselves as frequent daydreamers would express less aggression in the cathartic condition compared with subjects who were not frequent daydreamers. In addition, it was predicted that both the TAT fantasy and the daydreaming experience would be equally effective as cathartic conditions because the subject would be able to structure the content of the fantasy in each. Because no prior work was done on male-female differences in this area, no predictions were made regarding sex differences.

It was anticipated that the use of three dependent variables, each measuring aggressive feelings at different levels of awareness and towards different objects, would clarify the process of catharsis. Would, for example, the fantasy experience result in less aggressive feeling at the most overt attitudinal level, but not change a subject's self-reports of how aggressive a person he is, or change the even more covert aggression measured by a word-association test?²

² The third dependent variable, number of hostile words on the word-association test, could not be analyzed because one subgroup (low-daydream males under the insult control conditions) failed to produce the minimum number of associations to the stimulus word. This in itself seemed a noteworthy finding, as most subjects in other subgroups produced a large number of associations in the allotted time. In addition, a further shortcoming of this variable was its failure to discriminate insulted from noninsulted subjects, so it was consequently not possible to study catharsis in the absence of a measure sensitive to anger arousal.

METHOD

Experimental Design

A $2 \times 2 \times 2 \times 3$ analysis of variance design with a partitioning of 120 subjects into 24 subgroups was used. Four independent variables with their respective subgroups were: (a) sex (males and females), (b) daydreaming frequency (high and low), (c) insult conditions (insult and noninsult), (d) cathartic conditions (daydreaming, TAT, and control). The three dependent variables were: (a) direct verbal expression of hostility (attitude questionnaire), (b) self-ratings of hostility (Sarason Hostility scale), and (c) responsiveness to hostile stimuli (modified word-association test). Three separate analyses of variance were carried out, one for each dependent variable. In addition, t tests between individual subgroups were calculated where appropriate.

Two types of controls were employed. Because previous research has indicated that in the absence of drive arousal fantasy does not have a cathartic effect, a control group was needed to establish drive arousal. Therefore, to test the effectiveness of the anger-arousing condition, half of the subjects received the insulting instructions and half received noninsulting instructions. It was anticipated that insulted subjects would score higher on subsequent hostility scales than noninsulted subjects, thus establishing that drive arousal had occurred.

A second control group was needed to test the effectiveness of the two cathartic conditions. A control activity was designed which was neither stressful nor provided an opportunity for the displacement of aggression, but which would occupy the subjects in an interesting task for a time equivalent to the experimental subjects. It was anticipated that insulted subjects in the control condition would express considerably more subsequent hostility than subjects in either of the cathartic conditions.

Subjects

The subjects were 60 male and 60 female undergraduates selected on the basis of self-reports of daydreaming frequency from a population of 331 students in elementary psychology classes.

A daydreaming questionnaire devised by Singer was used (Singer & Antrobus, 1963), although the directions were modified (Pytkowicz, 1964). This questionnaire lists 120 daydreams to which subjects respond on 5-point frequency scale. A weighted score was obtained for each subject, representing the frequency with which he experiences this sample of daydreams. Singer reported that this score is related positively to information about daydreams obtained from interview data (Singer & Antrobus, 1963). Daydreaming scale mean scores were 121 for males and 109 for females, a difference which approaches significance ($p < .10$). Males and females from the upper and lower 25% of their respective distributions were selected as subjects.

Materials

Attitude questionnaire. This dealt with the subject's feelings about participating in psychological experiments and his evaluation of the experimenter and the experiment. It consisted of six questions, each with six alternative answers ranging from approval to strong disapproval. This questionnaire was used by Feshbach (1955, 1961) and was found to be sensitive to experimental conditions of anger arousal. By using a weighted scoring system allowing 1 point for a highly approving response and 6 points for strong disapproval, a total score was obtained for each subject, representing his expressed feelings of disapproval of the experiment. Thus, possible scores ranged from 6 (strong approval on all items) to 36 (strong disapproval on all items).

Sarason Hostility scale. This is a self-report questionnaire in which the subject indicates by responding T or F the extent to which he believes certain hostile descriptions apply to him (Ganzer & Sarason, 1964). Possible scores range from 0 to 23, with higher scores indicating self-ratings of greater hostility. In order to control for individual differences in level of hostility, two scores on this variable were obtained for each subject, one at the beginning of the quarter which was considered a pretest, and the second following the experimental conditions, called the posttest. The difference score (posttest minus pretest score) thus represented changes in the subject's self-ratings of hostility following the experimental conditions. A positive score represented a change in the direction of an increase in hostility ratings following the experiment, while negative difference scores indicated self-ratings of less hostility in the posttest situation.

Modified word-association test. This was devised by Gellerman (1956), used by Feshbach (1961), and is a measure found to be sensitive to differences in the arousal of aggression. However, in a recent study (Kaufmann & Feshbach, 1963) this measure did not discriminate insulted and noninsulted subjects. The subjects were instructed to write all associations to 11 stimulus words, presented for a period of either 1 minute (for the 6 neutral words) or 2 minutes (for the 5 hostile words). A hostility score was obtained from each subject on this measure, after the manner suggested by Gellerman, by totaling the number of hostile word associations given among the second set of five responses to each hostile stimulus word.

The instructions used in this study were somewhat different from those of Gellerman, who instructed that 10 words be written to each stimulus word. Because some subjects failed to write 10 associations, although given adequate time, and because the measure failed to discriminate between insulted and noninsulted subjects, this variable was not used in the study as planned. However, an interesting incidental finding was observed. Although some subjects in each subgroup failed to make 10 associations, the only subgroup in which all subjects

failed to respond was the insulted-low-daydream males in the control conditions. According to catharsis theory, this group, having been insulted and having had no opportunity for catharsis through fantasy, would have the highest level of residual hostility. It seems probable that this group, rather than expressing their hostility indirectly through a high level of hostile word associations, acted it out directly by refusing to comply with the directions and write as many associations as possible in the time allowed.

Procedure

At the beginning of the quarter, subjects were given the Autobiographical Survey (Sarason, 1958) which contained the Sarason Hostility scale (Ganzer & Sarason, 1964), a Test Anxiety scale, a General Anxiety scale (Sarason & Ganzer, 1962), and a Need for Achievement scale (Adams & Sarason, 1963). The Edwards (1957) Social Desirability scale was also administered. Two weeks before the experimental session, the entire psychology class was given the Singer Daydreaming Questionnaire. One hundred-twenty students were selected as subjects on the basis of their admitted frequency of daydreaming and were contacted by telephone and asked to participate in the experiment. They were randomly distributed into the six experimental conditions.

Experimental sessions were similar for each group of subjects, differing only in respect to the experimental variables. Sessions followed this plan: The study was introduced (with either insulting or non-insulting introductions), and the subjects were presented with booklets containing all the test materials. Cathartic conditions were then presented (groups had either a daydreaming activity, a TAT activity, or a control task). Following this, each group was given identical procedures, as each of the dependent variables was measured: modified word-association test, the Sarason Hostility scale, and the attitude questionnaire. The subjects were then given a brief account of the study and asked to maintain secrecy until all groups were tested. The entire procedure took 50 minutes.

Insult and noninsult conditions. Subjects in the insult conditions were subjected to a number of critical and unwarranted remarks following Feshbach (1955, 1961) and Gellerman (1956) to arouse aggressive drive. The subjects' academic motivation, emotional maturity, intelligence, and honor were challenged as the experimenter gave the introduction and directions. The experimenter's attitude was brusque and authoritarian, and she arrived a few minutes late after having insisted that the subjects arrive on time. Subjects in the noninsult conditions were greeted warmly, treated with respect, and given a neutral introduction.

Cathartic conditions. Subjects in the daydreaming condition were instructed to let their minds roam freely, making no effort at conscious control for a 10-minute period directly following the introduction. The experimenter sat quietly with the subjects

for 10 minutes, and then administered the three measures of hostility.

Subjects in the TAT condition were told a picture would be projected on the screen for 20 seconds and they were to write a story covering four general questions appearing in their answer booklets. Four TAT cards (4, 7BM, 11BM, and 18GF) were projected one at a time. This task took about 15 minutes and was introduced as a measure of imagination.

Subjects in the control condition were given a task described as a measure of "literary sensitivity." They were asked to read four selections taken from the works of famous authors and make judgments as to the age and sex of the author as well as telling the criteria they used in making their decisions.

RESULTS

Attitude Questionnaire

Table 1 presents the attitude questionnaire mean scores for the 24 subgroups. Analysis of variance of the 120 subjects' scores (Table 2) on the attitude questionnaire revealed significant main effects for three of the four variables: sex, insult conditions, and cathartic conditions. Males expressed significantly more hostility on the attitude questionnaire than females, insulted subjects expressed significantly more hostility than noninsulted subjects, and subjects in the daydream and control conditions expressed significantly more hostility than in the TAT condition. There was also a significant interaction between sex

TABLE 1
ATTITUDE QUESTIONNAIRE MEAN SCORES OF
INDIVIDUAL SUBGROUPS

Subgroups	Insult conditions	Noninsult conditions
High-Daydream males		
Daydream condition	16.8	14.0
TAT condition	16.0	15.2
Control condition	21.6	14.6
Low-Daydream males		
Daydream condition	19.6	17.8
TAT condition	19.0	14.2
Control condition	21.4	16.6
High-Daydream females		
Daydream condition	17.2	14.2
TAT condition	14.0	14.8
Control condition	16.0	14.2
Low-Daydream females		
Daydream condition	19.1	16.2
TAT condition	11.2	14.4
Control condition	15.2	15.0

Note.— $N = 5$.

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE OF
ATTITUDE QUESTIONNAIRE RESPONSES

Source of variation	SS	df	MS	F
Sex	134.40	1	134.40	9.57***
Daydreaming (DD)	25.20	1	25.20	1.79
Insult conditions (IC)	138.67	1	138.67	9.88***
Cathartic conditions (CC)	105.35	2	52.68	3.75*
Sex × DD	20.20	1	20.20	
Sex × IC	69.02	1	69.02	4.92*
Sex × CC	48.72	2	24.36	
DD × IC	2.41	1	2.41	
DD × CC	45.32	2	22.66	
IC × CC	49.55	2	24.76	
Sex × DD × IC	5.02	1	5.02	
DD × IC × CC	9.12	2	4.56	
Sex × IC × CC	48.61	2	24.31	
Sex × DD × CC	3.63	2	1.82	
Sex × DD × IC × CC	22.10	2	11.05	
Within	1348.00	96	14.04	
Total	2076.00	119		

* $p < .05$.*** $p < .01$.

and insult conditions, which reveals that the effects of the insult conditions apply only to males. There is no difference in the scores of insulted compared with noninsulted females (mean scores equal 15.4 and 14.8, respectively), while for males the scores of insulted compared with noninsulted subjects (mean scores of 19.01 and 15.4, respectively) are significantly greater. The insult conditions as measured by the attitude questionnaire were thus effective in arousing hostility in males but not in females. Therefore, in the absence of anger arousal in females, it is not possible to test for catharsis.

The cathartic effects of daydreaming and TAT experiences were tested by individual t tests for high- and low-daydreaming males. For insulted-high-daydream males, both daydreaming and the TAT conditions resulted in significantly fewer hostile responses to the attitude questionnaire compared with insulted control subjects ($t = 2.31, 2.21$, respectively; $p < .05$). Insulted-low-daydream males showed a similar, but nonsignificant trend. Thus, although both high- and low-daydream males expressed equivalent amounts of hostility under the insult control conditions, only the high-daydream males expressed significantly less hostile responses following the insult fantasy conditions. This finding supports one major prediction of this study, namely, that

males who describe themselves as frequent daydreamers will be able to utilize a fantasy experience more effectively for the reduction of aroused hostile feelings than will males who describe themselves as infrequently engaging in daydreaming. It is also of interest that both the TAT and the daydream conditions were equally effective cathartic experiences for insulted males.

Although an unfortunate random variation in the original (pretest) Hostility scale scores of the subjects in one particular subgroup (high-fantasy males in the insult control condition) might suggest that the best interpretation of the above findings would be in terms of regression to the mean, this does not seem a tenable hypothesis for the following reason: as Table 1 indicates, despite the initially higher pretest Hostility scale scores of high-fantasy male subjects in the insult control condition, it is noteworthy that their attitude scale scores following experimental conditions are identical to the low-fantasy males in the same experimental condition. This finding would suggest that experimental conditions rather than original level of hostility and regression to the mean represent the more parsimonious interpretation.

Hostility Scale

Reference to Table 3 reveals that, as predicted, the experimental conditions do not have the same effect on subjects' self-ratings of hostility as they do on hostile responses on the attitude questionnaire. An analysis of variance (Sex × Cathartic Conditions × Insult Conditions) of the Hostility scale difference scores revealed one significant main effect (cathartic conditions, $F = 6.66, p < .01$) and two significant interactions (Sex × Cathartic Conditions, $F = 5.79, p < .05$; Insult × Cathartic Conditions, $F = 7.18, p < .01$). All noninsulted subjects manifested a consistent increase in Hostility scale scores on the second administration of the scale (positive difference score). Likewise, insulted subjects under cathartic conditions also increased Hostility scale scores on the postexperimental administration, while insulted subjects having no opportunity for catharsis (controls) manifested significant decreases in Hostility scale

scores. As Table 3 also indicates, the difference between cathartic and control subjects is significant only for high-daydream subjects, both male and female, and, although the trends are in the same direction for low-fantasy subjects, the differences are not significant.

Thus, subjects (at least high-daydream subjects) in the insult control condition rated themselves as significantly *less* hostile on the second administration of the test compared to subjects in the insult cathartic conditions, who rated themselves as significantly *more* hostile on the second administration. This finding is somewhat difficult to interpret due to an unexpected random variation in the subjects' original Hostility scale scores, resulting in the high-daydream males in the insult control condition having higher scores than any of the other 23 subgroups. It would be possible to interpret this finding as a regression effect were it not for the fact that it is paralleled by the significant differences between high-daydream females in the insult control and cathartic conditions, subgroups in whom the initial hostility scores were all comparable.

Content Analysis of the TAT

The TAT stories of the 40 subjects in the TAT condition were scored for aggressive content to evaluate whether catharsis occurred through direct displacement of anger into fantasy. Two scales were developed for this study, one measuring direct aggression expressed in story content, and the other measuring indirect aggression. Each story was scored on a 5-point scale.

Direct aggression included any destructive behavior toward a person or his property by another person with accompanying anger or intent to harm, and included such behaviors as verbal anger, hitting, shooting, murder, etc. Indirect aggression meant aggression which was either not directly attributable to an individual (war, accidents, etc.), aggression not carried out with attempt to harm, or symbolic aggression (sickness, death from natural causes, etc.). As each subject had written one story to each of the four TAT cards, scores on each scale ranged from a possible 0 (no expressed aggression) to a score of 16.

A reliability study between two judges,

TABLE 3
HOSTILITY SCALE MEAN DIFFERENCE SCORES FOR INDIVIDUAL SUBGROUPS
AND SIGNIFICANCE TESTS FOR INSULT CONDITION ONLY

Subgroup	Noninsult conditions	Insult conditions		Comparison	<i>t</i>
	\bar{X}	\bar{X}	<i>s</i>		
High-Daydream males					
Daydream condition	2.8	3.7	3.09	Control & DD Control & TAT	3.173**
TAT condition	2.2	1.0	1.63		2.436*
Control condition	2.2	-3.5	3.32		
Low-Daydream males					
Daydream condition	4.5	3.0	3.16	Control & DD Control & TAT	1.484
TAT condition	1.2	0.0	2.16		<i>ns</i>
Control condition	4.0	-0.8	4.04		
High-Daydream females					
Daydream condition	0.5	3.7	0.5	Control & DD Control & TAT	9.270***
TAT condition	0.7	4.0	3.32		3.134**
Control condition	1.2	-1.5	1.0		
Low-Daydream females					
Daydream condition	1.7	2.0	2.45	Control & DD Control & TAT	<i>ns</i>
TAT condition	1.5	0.7	1.26		<i>ns</i>
Control condition	1.7	0.0	1.14		

Note.—Difference score is the pretest Hostility scale score subtracted from the posttest Hostility scale score. Magnitude of the difference score represents the amount of change in self-ratings of hostility. A positive difference score represents an increase in self-ratings of hostility, while a negative score represents a decrease in self-ratings of hostility from pretest to posttest. *N* = 4.

**p* < .05 (2-tailed test).

***p* < .02 (2-tailed test).

****p* < .01 (2-tailed test).

TABLE 4

PRODUCT MOMENT CORRELATIONS BETWEEN TAT AGGRESSION AND SELF-RATINGS OF HOSTILITY

TAT scores	Direct aggression	Indirect aggression
Attitude questionnaire		
Insulted Ss (<i>N</i> = 20)	.139	-.05
Noninsulted Ss (<i>N</i> = 20)	.018	-.164
Insulted males (<i>N</i> = 10)	.170	-.133
Sarason Hostility scale (original score)		
Insulted Ss (<i>N</i> = 19)	.538*	-.481
Noninsulted Ss (<i>N</i> = 17)	.366	-.452

**p* < .05 (2-tailed test).

each independently scoring the four stories of 10 subjects, resulted in a product-moment correlation of .986 for direct aggression and .929 for indirect aggression. Thus, the remainder of the stories were rated by the senior author.

There were no significant differences between the various groups of subjects in quantity or type of aggressive content appearing in their TAT stories. As Table 4 indicates, there was little relationship between the amount of hostility a subject expressed on the attitude questionnaire and either direct or indirect aggression expressed in TAT stories. It is also noteworthy that this lack of relationship remains essentially unchanged for subjects regardless of whether or not they have previously been insulted. These two findings suggest that although catharsis does occur through the fantasy experience, it occurs without the direct displacement of aggression into the fantasy content.

The further finding (Table 4) of a significant positive correlation between subjects' original hostility score on the self-rating scale and direct aggression in TAT stories, plus the fact that this correlation was significant only under insult conditions, suggest that the amount of aggression expressed in the TAT stories is a function of the subjects' personal level of hostility, enhanced by the experimental conditions of anger arousal.

This finding is compatible with a nonquantified impression of a general lack of direct displacement of aggression into daydreams. The subjects were asked at the end of the study to write a brief account of what it had felt like to daydream and what they had day-

dreamed about. The majority of these reported daydreams were either of fairly mundane content (tests, grades, dating, etc.) or were of a freely imaginative nature (flights into fantasy, imaginative conjecture, fragments from the past or future, etc.). Most subjects described the daydreaming period as relaxing or soothing, a few reported it was boring or difficult because it was not spontaneous, but no subject reported hostile or even anxious daydreams.

Summary of Results

For males, the effect of the insult compared with the noninsult conditions was to produce significantly more verbal expression of hostility (scores on an attitude questionnaire), and to significantly decrease self-ratings of hostility. For females, the insulting conditions did not produce differential responses on either of the above dependent variables, so in the absence of a measure of anger arousal it was not possible to study the effects of fantasy on anger reduction in females.

Insulted male subjects who themselves professed to frequent daydreaming had significantly lower hostility scores on an attitude questionnaire following fantasy activities than comparable insulted-high-daydream subjects in a control condition. Insulted-high-daydream males also rated themselves significantly more hostile following fantasy activities than did a comparable group of subjects following a control condition.

The above findings, stated differently, suggest that the cathartic effects of fantasy operate in different directions, depending on which measures of aggression are used. When the object of the aggression was the examiner (attitude questionnaire), fantasy following frustration resulted in lower aggression scores; when the object of the aggression was the self (Hostility scale), fantasy increased hostility scores. Although the fantasy experiences resulted in catharsis for both high- and low-daydream males, the fantasy condition had a more pronounced cathartic effect for the high-daydream males, as was predicted. The findings for low-daydream males followed a similar pattern, but did not reach significance.

In terms of the effectiveness of the two types of fantasy experience in producing catharsis, daydreaming, compared with the TAT condition, was as effective a cathartic experience for both high- and low-daydream males when aggression directed toward the examiner was measured. However, when aggression directed toward the self was measured, there was a trend in the direction of greater increases in the self-ratings of hostility following daydreaming than following the TAT. No simple relationship was found between the amount or type of aggression expressed in the TAT stories and either of the measures of aggression.

DISCUSSION

In general, the findings of this study support but modify catharsis theory, and delineate further the conditions under which the instigation to aggression can be reduced through a fantasy experience.

The study replicates the findings of Feshbach and others in demonstrating (for males) the cathartic effects of fantasy following frustration. Of further significance is the finding that although catharsis occurs under most conditions for most subjects, it is the men who report frequent daydreaming who can utilize a fantasy experience most effectively, for the reduction of aggressive feeling. This finding lends support to Singer's view of daydreaming as an organismic variable along which subjects vary and which can serve as an adaptive mechanism related to impulse control and self-awareness.

The findings of this study are in general not compatible with those of Berkowitz and Geen (1966) and other research employing film-mediated fantasy. Apparently the exciting, provocative, and aggressive films facilitate subsequent aggressive expression, while the more reflective and personalized fantasy of daydreams inhibits the direct expression of aggression. Further study of the type of fantasy experience related to catharsis is warranted. This study compared two types of fantasy and found them to be equally effective as cathartic experiences only when aggressive feelings toward the examiner were measured.

The discrepant findings between the two measures of aggression confirm the expectation that the presence or absence of catharsis depends, to some extent, on the measures used to assess aggression. It does not seem feasible to conceptualize "aggressive drive" independent of the methods used to measure it.

The complexity of these findings suggests that a hydraulic drive-reduction concept of catharsis is not sufficient. No simple displacement of aggressive affect into fantasy was found. Rather than being "discharged" into a substitute fantasy expression, the angry affect appears to be displaced onto the self following a fantasy experience. In this sense, fantasy following frustration appears to facilitate the internalization of anger. Further investigation is needed to ascertain if indeed this is the process by which fantasy effectively inhibits the more direct expression of affect.

REFERENCES

- ADAMS, E. B., & SARASON, I. G. Relation between anxiety in children and their parents. *Child Development*, 1963, **34**, 237-246.
- BERKOWITZ, L. Some factors affecting the reduction of overt hostility. *Journal of Abnormal and Social Psychology*, 1960, **60**, 14-21.
- BERKOWITZ, L., & GEEN, R. G. Film violence and the cue properties of available targets. *Journal of Personality and Social Psychology*, 1966, **3**, 525-530.
- DOLLARD, J., DOOB, L. W., MILLER, N. E., MOWRE, O. H., & SEARS, R. R. *Frustration and aggression*. New Haven: Yale University Press, 1939.
- DOLLARD, J., & MILLER, N. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
- EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.
- ESTESS, B. The cathartic effect of 2 kinds of phantasy aggression on aggressive behavior. Unpublished honor thesis, Smith College, 1960.
- FESHBACH, S. Drive reducing function of fantasy behavior. *Journal of Abnormal and Social Psychology*, 1955, **50**, 3-11.
- FESHBACH, S. The catharsis hypothesis and some consequences of interaction with aggression and neutral play objects. *Journal of Personality*, 1956, **24**, 449-462.
- FESHBACH, S. The stimulating versus cathartic effects of a vicarious aggressive activity. *Journal of Abnormal and Social Psychology*, 1961, **63**, 381-385.
- FREUD, S. Beyond the pleasure principle. *The complete psychological works of Sigmund Freud*. Vol. 18. London: Hogarth, 1920. Pp. 3-64.

- FEUD, S. The relationship of the poet to daydreaming. *Collected papers*. Vol. 4. London: Hogarth, 1925. Pp. 173-183.
- GANZER, V. J., & SARASON, I. G. Interrelationships among hostility, experimental conditions, and verbal behavior. *Journal of Abnormal and Social Psychology*, 1964, 68, 79-84.
- GELLERMAN, S. The effects of experimentally induced aggression on inhibition on word association response sequences. Unpublished doctoral dissertation, University of Pennsylvania, 1956.
- GORDON, J., & COHN, F. Effect of fantasy arousal of affiliation drive on doll play aggression. *Journal of Abnormal and Social Psychology*, 1963, 66, 301-307.
- KAUFMAN, H., & FESHBACH, S. Displaced aggression and its modification through exposure to anti-aggressive communication. *Journal of Abnormal and Social Psychology*, 1963, 67, 79-83.
- KENNY, D. T. An experimental test of the catharsis theory of aggression. Unpublished doctoral dissertation, University of Washington, 1952.
- MUSSEN, P., & RUTHERFORD, E. Effects of aggressive cartoons on children's aggressive play. *Journal of Abnormal and Social Psychology*, 1961, 62, 461-464.
- PYTKOWICZ, A. R. An experimental study of the relationship of fantasy to the reduction of hostility. Unpublished doctoral dissertation, University of Washington, 1964.
- SARASON, I. G. Interrelations among individual difference variables, behavior in psychotherapy and verbal conditioning. *Journal of Abnormal and Social Psychology*, 1958, 56, 339-344.
- SARASON, I. G., & GANZER, V. J. Anxiety, reinforcement, and experimental instructions in a free verbalization situation. *Journal of Abnormal and Social Psychology*, 1962, 65, 300-307.
- SAXTON, G. Spontaneous fantasy as a resource of high grade retardates for coping with a failure-stress frustration. *Journal of Abnormal and Social Psychology*, 1962, 64, 81-84.
- SIEGEL, A. Film-mediated phantasy aggression and strength of aggressive drive. *Child Development*, 1956, 27, 365-378.
- SINGER, J. L. Delayed gratification and ego-development: Implications for clinical and experimental research. *Journal of Consulting Psychology*, 1955, 19, 259-266.
- SINGER, J. L. Imagination and waiting ability in young children. *Journal of Personality*, 1961, 29, 396-413.
- SINGER, J. L. *Daydreaming: An introduction to the experimental study of inner experience*. New York: Random House, 1966.
- SINGER, J. L., & ANTROBUS, J. S. A factor-analytic study of daydreaming and conceptually-related cognitive and personality variables. *Perceptual and Motor Skills*, 1963, 17, 187-209.
- SINGER, J. L., & MCCRAVEN, V. Some characteristics of adult daydreaming. *Journal of Psychology*, 1961, 51, 151-164.
- SINGER, J. L., & MCCRAVEN, V. Patterns of daydreaming in American subcultural groups. *International Journal of Social Psychiatry*, 1962, 8, 272-282.
- SINGER, J. L., & OPLER, M. Contrasting patterns of phantasy and motility in Irish and Italian schizophrenics. *Journal of Abnormal and Social Psychology*, 1956, 53, 42-47.
- SINGER, J. L., & ROWE, R. An experimental study of some relationships between daydreaming and anxiety. *Journal of Consulting Psychology*, 1962, 26, 446-454.
- SINGER, J. L., & SCHONBAR, R. A. Correlates of daydreaming: A dimension of self-awareness. *Journal of Consulting Psychology*, 1961, 25, 1-6.
- SINGER, J. L., WILENSKY, H., & MCCRAVEN, V. Delaying capacity, phantasy and planning abilities: A factorial design of some basic ego functions. *Journal of Consulting Psychology*, 1956, 20, 375-383.

(Received May 2, 1966)

RESPONSE AVAILABILITY AS A FACTOR IN THE PROBLEM-SOLVING PERFORMANCE OF MALES AND FEMALES¹

NORMAN R. F. MAIER AND RONALD J. BURKE

University of Michigan

314 male and 370 female college students were given 1 of 4 versions of the horse-trading problem or 1 of 2 versions of the used-car problem (a similar problem but with no correct or incorrect responses) in a series of experiments investigating sources of difficulty within problems. 2 alternatives which distracted Ss from the correct answer to the horse-trading problem were found. 1 applied to both males and females and seemed to be the result of a faulty perception of the problem. The other was a generalized tendency for females to choose the broke-even alternative when money transactions were involved. Significant differences ($p < .01$) between correct responses for males and females were found whenever this choice was available and was not the correct one. When the problem was altered so that it was the correct answer, females performed about as well as males. A secondary finding was that on the used-car problem females avoided the lost-money response when a female was seen to be the cause of the loss, but not when the sex roles were reversed. Comparisons of problem-solving performance of males and females can lead to ambiguous conclusions unless their response repertoires are equated. If incorrect alternatives occupy different positions in the response hierarchies of the 2 sexes, the problem used cannot be a comparable test of problem-solving ability or motivation.

Failure to solve a problem may have a variety of causes, and sometimes more can be learned from the study of failures than of successes. In the comparative study of problem solving in animals it was found necessary to locate the sources of difficulty within problems. These sources depend on the behavior repertoires the various organisms bring with them to the problem situation. The task of learning to pull a string to open a food box, for example, is easier for cats than for dogs, because cats are the more likely to react to a dangling string. Similarly, a pig is more likely to solve a problem that requires the raising of a small platform than is a dog because of the rooting tendency in pigs. Cats are easier to housebreak than monkeys because one can capitalize on responses displayed before training. It would be an error to judge one species to be more intelligent than another because it learned or solved such problems the more readily. This is true

largely because the experimenter can manipulate the outcome of such comparisons by requiring responses that are more available in one animal form than in another.

Native endowments in the behavior repertoire an organism brings to the situation, therefore, can influence whether a task assignment is easy or difficult for a given species. In addition, previous experiences increase the repertoire and make certain responses more available than others. If a child is given a stick and is asked to get a banana from a bunch hanging overhead, he uses the stick to knock one off, but a chimpanzee given the same problem climbs up the stick and grabs a banana with his hand. Previous experience has given the stick different meanings or functions for them, and it would be inappropriate to assume that one solution was superior to the other. It would be misleading to expect the chimpanzee to solve the problem by a method that presupposed certain specialized knowledge available only to the child. Similarly, problems requiring a knowledge of calculus are inappropriate for the study of problem solving unless all subjects have had

¹ The research reported here was supported by Grant No. MH-02704 to N. R. F. Maier from the United States Public Health Service.

We wish to gratefully acknowledge the assistance of Junie Janzen in this study.

adequate training. Simple arithmetic problems often are used in such studies because it is assumed that the essential skill requirements have been learned and are readily available to all subjects.

However, is this a fair assumption? If native differences or certain past experiences serve as an aid to problem solving, it must also be recognized that they can be a handicap. Thus the chimpanzee's experience with climbing may cause him to fail to use a stick as a way to extend his reach. The tendency to reach incorrect solutions, therefore, may be caused not only by deficiencies in thinking, but also by distracting past experiences.

It is common knowledge that multiple-choice questions can be made more difficult either by (a) altering the stem or the correct alternative, or (b) making the incorrect alternatives more plausible or more difficult to distinguish from the correct one. Plausible incorrect alternatives serve as distractors and hence become sources of error.

Perhaps the problem-solving behavior of males and females should be reexamined in the light of these implications. Hoffman and Maier (1966) found it difficult to explain sex differences in problem solving in terms of differing abilities, strength of motivation, different reactions to the sex of the experimenter, or type of problem content (such as masculine versus feminine versions of the problem), because none of these variables could be generalized. Each type of problem seemed to require a different explanation to account for the relatively poorer performance of females.

Is it possible that females do not do as well because certain incorrect alternatives have a stronger attraction for them? To test this possibility comparisons of all responses rather than of only the correct responses must be made. For example, females may not do as well as males on a particular problem because an incorrect alternative is a relatively greater distractor for them. If this is true, then the same objective problem is different and more difficult when given to females than when given to males. Such a problem, therefore, would not permit a valid comparison of their problem-solving abilities nor their motivations. The purpose of the present investiga-

tion is to test the hypothesis that the hierarchy of responses differs for male and female subjects. In other words, the competition between response alternatives is not the same for the sexes, and this may account for some of the differences in their performance in problem situations.

General Procedure

Four variations of the horse-trading problem (Maier & Solem, 1952) and two versions of a new problem, called the used-car problem, were used. These six conditions will be described in sequence. In each condition male and female subjects were obtained from our subject pool (made up of students in introductory psychology courses).

EXPERIMENT I

Procedure

The standard version of the horse-trading problem was used. Each student received a sheet of paper on which the problem was stated as follows: "A man bought a horse for \$60 and sold it for \$70. Then he bought it back again for \$80 and sold it for \$90. How much money did he make in the horse business?" Five alternative answers were listed (lost \$10, broke even, made \$10, made \$20, made \$30), and the subject was asked to check one answer only. (Previous research has indicated that all of these answers are given when the alternatives are not supplied.) A total of 87 males and 86 females participated in this experiment.

Results

The results of this experiment are shown in Table 1. The correct answer (made \$20) was chosen by 48.3% of the males and by

TABLE 1
FREQUENCY OF MALE AND FEMALE ANSWERS
TO STANDARD HORSE-TRADING PROBLEM

Solutions offered	% selecting each solution	
	Males ($N = 87$)	Females ($N = 86$)
Lost \$10	1.1	1.2
Broke even	9.2**	27.9**
Made \$10	37.9	40.9
Made \$20	48.3**	26.7**
Made \$30	3.4	3.5

** $p < .01$ for difference between males and females.

26.7% of the females, a difference significant at the .01 level. Note, however, that the alternative made \$10 was chosen with about equal frequencies by the two sexes, while alternative broke even was chosen more frequently by the females. Can this latter alternative be a unique distractor for females? If the percentages of those choosing the alternatives broke even and made \$20 are combined, the males' and females' scores are very similar, being 57.5% and 54.6%, respectively.

Perhaps females are less inclined to think things through and are prone to jump to conclusions. The following experiment was performed to test this possibility.

EXPERIMENT II

Procedure

This experiment was the same as Experiment I except that a rationale for each alternative answer was added. This rationale was intended to cause the subject to examine alternatives and therefore reduce the tendency to jump to a quick conclusion. The following form, with the several rationales supplied, is shown below.

A man bought a horse for \$60 and sold it again for \$70. Then he bought it back again for \$80 and sold it for \$90. How much money did he make in the horse business?

Read all five alternatives before selecting the one you think is correct. Check one answer only.

_____ He lost \$10. He lost \$20 by paying \$80 for a horse which he originally bought for \$60, but he made \$10 by selling it again for \$90. The sum of his gains and losses still yields a loss of \$10.

_____ He broke even. As a result of the first sale the trader had \$70. He then borrowed \$10 to be able to buy the horse back for \$80. The \$10 profit he made on the second sale was then needed to pay back the \$10 he borrowed and thus he broke even.

_____ He made \$10. He made \$10 by buying the horse for \$60 and selling it for \$70. He lost \$10 by buying the horse back for \$80, but then made \$10 by selling it again for \$90. The sum of his gains and losses yields a profit of \$10.

_____ He made \$20. He made \$10 by buying the horse for \$60 and selling it for \$70. He made another \$10 by buying it for \$80 and selling the horse again for \$90. The total profit is \$20.

_____ He made \$30. He started with \$60 and ended with \$90. The difference is \$30, his profit in the horse business.

A total of 43 males and 83 females were used in this experiment.

Results

The results of this experiment are shown in Table 2. We now find an increase in the number of correct responses for both males and females. The male increase (from 48% to 67.4%) is significant at the .01 level while the increase for the females (26.7% to 33.7%) is not significant. Supplying the rationale for the responses has helped the females less than the males, and hence it cannot be argued that the poorer performance of females is due to a tendency to jump to conclusions.

Of interest is the fact that females choose the broke-even solution with about the same frequency as in Experiment I. If we again combine the broke-even alternative with the correct answer, we find the score of females approaching that of males, the totals being 60.2% and 69.7%, respectively. This difference ceases to be significant.

If females choose the correct answer less frequently than males because the broke-even alternative has high availability, what will happen if this alternative is not offered?

EXPERIMENT III

Procedure

In this experiment the horse-trading problem was presented as in Experiment I, but the broke-even alternative was eliminated. The subjects consisted of 42 males and 72 females.

Results

Table 3 shows the results of this experiment. With the broke-even alternative elimi-

TABLE 2

RESPONSES OBTAINED WHEN RATIONALE TO THE HORSE-TRADING PROBLEM IS SUPPLIED

Solutions offered	% selecting each solution	
	Males (N = 43)	Females (N = 83)
Lost \$10	0.0	1.2
Broke even	2.3**	26.5**
Made \$10	27.9	34.9
Made \$20	67.4**	33.7**
Made \$30	2.3	3.6

** $p < .01$ for difference between males and females.

TABLE 3

PROBLEM CONCERNING A Horse-Trading Problem
WITH THE CORRECT ANSWER ALTERNATIVE OF BROKE-EVEN

Solutions offered	% selecting each solution	
	Males (N = 42)	Females (N = 50)
Lost \$20	2.4	0.0
Made \$20	47.6	44.0
Broke even	47.6	47.2
Made \$80	2.4	6.9

nated the results of males and females become almost identical and very much like that of the male group in Experiment I.

It appears that the horse-trading problem is more difficult for females because the broke-even answer is more of a distraction for them than for males. Does this tendency apply to this particular problem only or is this response more available in other problems involving money?

EXPERIMENT IV

Procedure

In this particular problem the subjects were given the problem form shown below.

A man bought a second hand car for his wife. She didn't like it so he sold it. How did he make out financially? One of the following alternatives is correct. Please choose one of them.

- _____ He lost money.
- _____ He broke even.
- _____ He made a small profit.
- _____ He made a good profit.

A total of 64 male and 50 female subjects were used.

Results

The results of this experiment are shown in Table 4. When no financial figures are given

TABLE 4

RESPONSES TO USED-CAR PROBLEM
(NO CORRECT ANSWER POSSIBLE)

Solutions offered	% selecting each solution	
	Males (N = 64)	Females (N = 50)
Lost money	75.0**	36.0**
Broke even	6.2**	28.0**
Made a small profit	9.4	22.0
Made a good profit	9.4	14.0

** $p < .01$ for difference between males and females.

to influence responses in a particular direction, we find male and female responses differing; in two of the four alternatives the differences are significant. As expected, females significantly favor the broke-even alternative more than males (28.0% versus 6.2%), and it is the highest availability of this alternative that seems to cause women to make the greater number of errors on the horse-trading problem.

However, women in this problem also choose the lost-money alternative significantly less often than men. On the basis of this difference it might be argued that women identify with the problem and feel guilty for not liking the car. Thus their choice of broke-even as well as their reduced preference for lost money suggest that our results may be influenced by these protective responses. A further experiment is needed before we can regard the broke-even response more highly available to females than to males.

EXPERIMENT V

Procedure

To test whether the choice of broke-even in the previous problem was an artifact of the husband-wife relationship the role functions were reversed and the form shown below was used.

A woman bought a second hand car, but her husband didn't like it. So a few days later she sold it. How did she make out financially? One of the following is correct. Please choose one of them.

- _____ She lost money
- _____ She broke even
- _____ She made a small profit
- _____ She made a good profit

TABLE 5

RESPONSES TO USED-CAR PROBLEM WITH
ROLES REVERSED

Solutions offered	% selecting each solution	
	Males (N = 34)	Females (N = 35)
Lost money	82.4*	57.1*
Broke even	5.9**	31.4**
Made a small profit	11.8	11.4
Made a good profit	0.0	0.0

Note.—For the percentage of females selecting the lost-money solution, the difference between corresponding items in Tables 4 and 5 is significant at .05.

* $p < .05$ for difference between males and females.

** $p < .01$ for difference between males and females.

The subjects consisted of 34 male and 35 female students.

Results

The results of this experiment are shown in Table 5. The broke-even category again is chosen significantly more often by females than by males, and the figures are approximately the same as in Table 4. Thus the reversed-sex version of this problem has not influenced the broke-even category.

Comparison of the lost-money categories of Tables 4 and 5 indicates no significant change in male responses, but female responses do differ. In Table 5 this category was chosen by 57.1% of the females as compared to 36.0% in Table 4. This difference is significant at the .05 level. Thus the protective response of the female disappears when she conducts the sale of the car rather than is the cause of it.

EXPERIMENT VI

Thus far all of the evidence indicates that females compared to males have a rather generalized tendency to favor the broke-even response. All differences were found to be significant at the .01 level of confidence. If our analysis is sound, it follows that if broke even were made the correct answer, women would no longer be at a disadvantage. Such a problem would be equivalent to combining broke even and correct responses, as was done in Experiments I and II.

Procedure

The horse-trading problem was modified so as to make broke even correct. The problem was formulated as given below.

A man bought a horse for \$60 and sold it for \$70. Then he bought it back for \$90 and sold it for \$80. How much money did he make in the horse business?

- _____ He lost \$20.
- _____ He lost \$10.
- _____ He broke even.
- _____ He made \$10.
- _____ He made \$20.

It should be noted that the initial amount of money and the final amount are made to differ so as to avoid a naive broke-even response.

The subjects consisted of 44 males and 44 females.

TABLE 6

FREQUENCY OF ANSWERS TO MODIFIED
HORSE-TRADING PROBLEM

Solutions offered	% selecting each solution	
	Males ($N = 44$)	Females ($N = 44$)
Lost \$20	6.8	15.9
Lost \$10	9.1	6.8
Broke even	81.8	70.5
Made \$10	0.0	6.8
Made \$20	2.3	0.0

Results

The results of this version of the horse-trading problem are shown in Table 6. We now find the correct responses (broke even) of males to be slightly higher than that of females, but for the first time the difference does not approach significance. In making the broke-even response and the correct response the same, the sex difference obtained up to this point disappears. Thus, either eliminating the broke-even response (Experiment III) or making it the correct response tends to equalize male and female performance on this type of problem.

DISCUSSION

Solving the horse-trading problem merely requires the proper application of arithmetic to some simple transactions. Unlike a creative problem, which involves the generation of ideas, the solution to the problem used in this investigation demands only that a correct procedure be followed. Since all of our subjects obviously possessed the essential skills of addition and subtraction, what makes the problem difficult? The analysis of the incorrect responses reveals that there are two misleading distractors. One of these applies to all subjects, the other is associated with females.

Let us consider first the distractor that misleads both males and females. Made \$10 is the most common incorrect answer for both sexes and in noncollege groups is the majority answer. This response seems to be a result of perceiving three separate financial transactions. Buying the horse for \$60 and selling

it for \$70 yields a \$10 profit; selling it for \$70 and buying it back for \$80 causes a \$10 loss; and finally buying it for \$80 and selling for \$90 yields another \$10 profit. The net result of the three transactions is a gain of \$10.

The problem becomes a simple one when viewed as two, rather than three, transactions. This correct approach can be induced readily in subjects simply by stating the problem as follows: "A man bought a white horse for \$60 and sold it for \$70. Then he bought a black horse for \$80 and sold it for \$90. How much did he make in the horse business?" The correct response of made \$20 is seldom missed in this version of the problem. We therefore may regard the two transactions with one horse as a more difficult problem than two transactions with different horses, even though the arithmetic skill requirements are the same. The horse-trading problem is difficult not because the answer is hard to figure out, but because an incorrect perception serves as a distractor. This distractor traps both males and females.

In order to explain the sex difference we must explore the second distractor. Analysis of our experimental variations showed the broke-even response to be a stronger distractor for females than for males. Significant differences were obtained whenever this choice was available and was not the correct answer. This specific difference in male and female responses is the basis for our contention that this problem is a more difficult one for females than for males and hence ceases to be a comparable test of problem-solving performance. Both males and females are subject to distractors, but in this instance the females have an added distractor, and hence problems involving such differences are not measures of problem-solving ability but tests of differences in behavior repertoires.

The reasons for such differences in behavior repertoires pose quite a different question. Why should women have a stronger tendency to have a business transaction *break even*? Cultural factors do influence response tendencies; we found a difference between Experiments IV and V when we reversed the sex roles in the used-car problem, which quite

obviously was caused by differences in the way male and female subjects perceived the male and female roles. In this instance the female selected the lost-money category less often when she did not like the car than when her husband did not like it. Female identification apparently caused females to defend themselves from being found to be a cause of a loss in money. However, this analysis does not explain the female attraction for the broke-even distractor, which was evident in our experimental series.

The preference for the broke-even response seems to be more general and not so dependent on the content of the problem. Our impression is that women are more conservative and protective. Taking chances on winning or losing are aggressive acts and these usually are associated with male personalities. Exploration of the literature supports these impressions. It has been shown that females are less aggressive (Berkowitz, 1962, pp. 267-274), more conforming (Kretch, Crutchfield, & Ballachey, 1962, pp. 522-525), and less willing to take extreme risks in gambling situations than males (Coombs & Pruitt, 1960; Kass, 1964).

Whether these differences in the personality traits of males and females are genetically or culturally determined remains unanswered. It is clear, however, that simple answers, such as male and female versions of the story, differences in motivation, women's lack of interest, etc., are rather superficial explanations of sex differences in problem-solving performance. Although such factors partially explain some of the experimental results, they cannot deal effectively with differences such as were obtained in the present series of experiments. The behavior repertoire differs for males and females, and these basic differences need to be explored before we appraise motivational and cultural variations.

REFERENCES

- BERKOWITZ, L. *Aggression: A social psychological analysis*. New York: McGraw-Hill, 1962.
- COOMBS, C. H., & PRUITT, D. G. Components of risk in decision making: Probability and variance pref-

- erences. *Journal of Experimental Psychology*, 1960, 60, 265-277.
- HOFFMAN, L. R. & MAIER, N. R. F. Social factors influencing problem solving in women. *Journal of Personality and Social Psychology*, 1966, 4, 382-391.
- KASS, N. Risk in decision making as a function of age, sex, and probability preferences. *Child Development*, 1964, 35, 577-582.
- KRETCH, D., CRUTCHFIELD, R. S., & BALLACHEY, E. L. *Individual in society*. New York: McGraw-Hill, 1962.
- MAIER, N. R. F., & SOLEM, A. R. The contribution of a discussion leader to the quality of group thinking: The effective use of minority opinions. *Human Relations*, 1952, 5, 277-288.

(Received May 23, 1966)

Manuscripts Accepted for Publication in the
Journal of Personality and Social Psychology

- Approach and Avoidance: Alternate Modes of Selective Exposure to Information: Aaron Lowin*: Department of Psychology, Iowa State University, Ames, Iowa 50010.
- Birth Order and Dependency: Ralph H. Stewart*: California State College, 5151 State College Drive, Los Angeles, California 90032. *
- Defensiveness in Decision Making as a Function of Sex-Role Identification: Paul D. Lipsitt* and Fred L. Strodtbeck: Law-Medicine Institute, Boston University, 141 Bay State Road, Boston, Massachusetts 02215.
- Awareness, Learning, and the Beneficent Subject as Expert Witness: Leon H. Levy*: Psychological Clinic, Indiana University, Bloomington, Indiana 47405.
- Odor as a Facilitator of the Effects of Subliminal Stimulation: Walter Gadlin and Harry Fiss*: Research Center for Mental Health, New York University, 4 Washington Place, New York, New York 10003.
- The Achievement Motive and Advancement in Two Types of Organizations: John Andrews*: Department of Psychology, Brandeis University, Waltham, Massachusetts 02154.
- Effects of Parental Dominance, Warmth, and Conflict on Imitation in Children: E. Mavis Hetherington* and Gary Frankie: Department of Psychology, University of Wisconsin, Madison, Wisconsin 53706.
- Effect of Race of Experimenter, Instructions, and Comparison Population upon Level of Reported Anxiety in Negro Subjects: Stephen S. Baratz*: Department of Psychology, Howard University, Washington, D. C. 20001.
- No Play: A Means of Conflict Resolution: Ralph R. Miller*: Department of Psychology, Rutgers—The State University, New Brunswick, New Jersey 08903.
- Generalization and Awareness in Verbal Operant Conditioning: Falak Thaver and William F. Oakes*: Department of Psychology, University of Hawaii, 2430 Campus Road, Honolulu, Hawaii 96822.
- Energizing (Dynamogenic) Effect of Cognitive Dissonance on Task Performance: Caroline K. Waterman* and Edward S. Katkin: Department of Psychology, State University of New York, Buffalo, New York 14214.
- Conformity as Related to Reinforcement and Social Pressure: Norman S. Endler* and Elizabeth Hoy: Department of Psychology, York University, Toronto 12, Ontario, Canada.
- Achievement and Affiliation Imagery in American Fiction: 1901-1961: Nancy M. Henley*: Department of Psychology, Johns Hopkins University, Baltimore, Maryland 21218.
- When Does a Harm Doer Compensate a Victim?: Ellen Berscheid* and Elaine Walster: Laboratory for Research in Social Relations, University of Minnesota, Minneapolis, Minnesota 55455.
- Comparison of Individual and Group Performance of Size Three and Five with Various Initially Right and Wrong Tendencies: Morton Goldman*, Ana McGlynn, and Ann Toledo: Psychology Department, University of Missouri, Kansas City, Missouri 64110.
- Aging among the Highland Maya: A Comparative Study: David Gutmann*: Department of Psychology, University of Michigan, Ann Arbor, Michigan 48104.
- Information Search as a Function of Conceptual Structure in a Complex Problem-Solving Task: Marvin Karkins* and Helmut Lamm: Green Hall, Department of Psychology, Princeton University, Princeton, New Jersey 08540.
- Language Variables Affecting the Persuasiveness of Simple Communications: David E. Kanouse and Robert P. Abelson*: Department of Psychology, Yale University, New Haven, Connecticut.

* Asterisks indicate author for whom address is supplied.

TENDENCIES TO COMPETE AND TO ATTACK AS A FUNCTION OF INSPECTION, INCENTIVE, AND AVAILABLE ALTERNATIVES¹

HARVEY A. HORNSTEIN AND MORTON DEUTSCH

Teachers College, Columbia University

Pairs of Ss played an experimental game designed to investigate some of the determinants of the tendencies to engage in cooperative, individualistic, and competitive modes of behavior. Results indicate that the allocation of effort among these alternative modes is associated with the incentive value of the alternative. In addition, there is evidence suggesting that the introduction of a cooperative alternative will reduce competitive behavior proportionally more than individualistic behavior. Further, other findings suggest that as the perceived likelihood of being inspected increases the readiness to engage in aggressive behavior may decrease. Moreover, providing players with the opportunity to inspect each other without requiring inspection had no noticeable advantage over a no-inspection condition. Requiring Ss to inspect one another, however, produced higher joint profits. Other findings indicate that the more aggressive member of the dyad generally profits more than the other player, but Ss in the more aggressive dyads generally earn less than Ss in the less aggressive dyads. Moreover, aggression results in relatively lower joint profits when the payoff structure favors cooperation rather than competition.

The present experiment is part of a program of studies concerned with the investigation of the determinants of the tendencies to engage in cooperative, individualistic, or competitive modes of behavior. In this study, the research is focused upon the effects of three variables: the availability of a mode of cooperative behavior, the relative strengths of the incentives for engaging in the different kinds of behavior, and the availability of different types of methods for obtaining information about the activities of the other person (inspection procedures).

Contrary to the postulate of "the independence of irrelevant alternatives" (Luce & Raiffa, 1958), which is one of the assumptions underlying some mathematical treatments of utility theory and of bargaining theory, one can assume that a person's preference for K (competitive activity) over I (individualistic activity) may be influenced by the presence or absence of the opportunity to choose C (cooperative activity). More specifically, it is suggested that the ratio of K choices to I choices is likely to be decreased if a person is given the opportunity

of choosing among K, I, and C, compared to when he is limited to choosing between K and I. We assume that introducing a cooperative alternative will stimulate cooperative behavior and, in so doing, will reduce the occurrence of both K and I—reducing K more, proportionally, than I. The notion here is that the processes underlying C and K are more mutually incompatible than those underlying C and I.

The assumption with regard to the effects of the strength of the economic incentives for engaging in the different behaviors is a traditional one: the stronger the incentive for engaging in a given mode of behavior, the more likely it is to occur. However, we do not suppose that economic incentives are the only incentives at work; personality and situational determinants may also affect the value of a given mode of behavior. Moreover, in a situation of uncertainty—when the objective probability that a given mode of behavior will lead to a given outcome is not specifiable—the likelihood of choosing any given mode is affected by the subjective estimate of the choices of obtaining the outcome associated with that mode. Thus, even if cooperation is seen to be of great potential reward, it is unlikely to be chosen unless one

¹ This research was supported by the National Science Foundation Grant GS-302, whose principal investigator is Morton Deutsch.

anticipates that one's choice to cooperate will be reciprocated, permitting cooperation to be realized.

In the present experiment, the inspection procedures provided completely accurate and trustworthy information, and the subjects believed this to be the case. Under such conditions, one can assume that the availability of an inspection procedure inhibits competitive behavior and facilitates cooperative behavior. The authors' reasoning is that since surprise was essential if one were to gain from competitive activities (in the game employed in the experiment), then the possibility of inspection would deter competition. On the other hand, it was hypothesized that cooperative activity would be more likely if the subjects were able to perceive that cooperation is mutual and inspection could permit the verification of its mutuality.

Three different inspection systems were studied: "mandatory," "voluntary symmetric," and "voluntary asymmetric." A fourth condition, "no inspection," did not allow inspection. The three inspection systems are described below in the Method section. Initially, there were no clear notions which would lead one to assume that the different system would have unique effects. However, it was expected that competition would be more inhibited the greater the number of inspections. In addition, since inspection would facilitate cooperation and inhibit competition, it was assumed that the stronger the incentive, the more likely subjects would use it to engage in cooperative activity. On the other hand, if the incentive to competitive activity were strong, they would be less likely to use inspection.

METHOD

In this game, the subjects had to allocate their efforts to the making of one or more of three different kinds of products.² They had 20 minutes during which they were permitted to make as many or as few of each type as they could. In order to produce these products they had to color two squares ($1\frac{1}{2} \times 1\frac{1}{2}$ inches) that were preprinted

on sheets of paper. Different colors were used to represent the different products.

Black products (I). If a player produced a black product, he received a specified amount of money for each one that he produced. Since the production of black squares did not affect the other player's profits, and since the value of the black squares to the subject remained unaffected by the other player's production, this mode of behavior was labeled "individualistic."

Blue products (C). In contrast, the value of a blue square was affected by the production of the other player. When both players made blue squares, each of their squares was worth a relatively high amount of money; if, however, a player's blue square was unmatched by the other player, it had a considerably smaller value. Thus, the production of blue squares was worthwhile only if the two players were able to cooperate; this is why it was labeled as "cooperative."

Red products (K). These products, referred to as "weapons" in this paper, had no value in and of themselves, but could be used to capture wealth from the other player. Each player had the power to attack in an attempt to gain money from the other player. In an attack, the weapons of the two players neutralized one another so that neither gained if both had the same number. However, if a player (whether he was the attacker or the attacked) neutralized all the weapons of the other player and had some additional weapons, he won a certain amount of money from the other player for each remaining weapon. In other words, no matter who attacked, the player with the greater number of weapons received money from the other, the amount of money being a function of his numerical superiority in weapon production. After an attack each player was left without any weapons; hence, the profit or loss in the next attack depended on the number of red squares produced since the previous one. The production of red squares was labeled as "competitive."

Inspection Procedures

Superimposed over this basic task were four different conditions involving the use of inspection: (a) *No inspection*—the game was played as described above, and the subject had no opportunity to find out what the other had produced; (b) *voluntary symmetric*—in this case, each subject had the right to ask for an inspection of the other's production to that point in the game. He was charged \$.06 each time he received this information; (c) *mandatory inspection*—three times during the course of the 20 minutes (after 6, 12, and 18 minutes) the experimenter told each of the subjects what the other had produce to that point; (d) *voluntary asymmetric*—only one subject had the right to request an inspection. As before, this information cost him \$.06. In all conditions, a subject knew when his activity was being inspected by the other subject.

Prior to the onset of the task, the subjects were

² This game was developed by Deutsch and Hornstein, collaboratively; it has in further experiments been modified so as to distinguish between aggressive and defensive forms of competitive behavior and to permit altruistic behavior.

fully informed about the nature of inspection procedures, the values and usages of the different products, and the mechanics of attacks. The values for I, C, and K were systematically varied in order to examine the effect of economic incentives on choice. As a result there were nine experimental treatments, a summary of which appear in Table 1.

In this experiment, 182 undergraduate or graduate students attending a summer session were randomly paired and assigned the nine experimental treatments. The paired subjects did not know each other's identity, nor were they allowed to communicate during the game. All subjects were told to earn as much as they could for themselves regardless of how much the other earned. Each subject was given \$.30 before the play began. They understood that they would keep whatever they won during the experiment.

Results

The best single measure of the players' strategies was the competition-noncompetition (CNC) index, which was computed as follows $(K - C - I)$ divided by $(K + C + I)$. When a player produced only K, the index was equal to 1.00, and when he produced only C or I it is equal to -1.00. If he produced an exactly equal amount of K and non-K $(C + I)$, the index was equal to zero.

Availability of a cooperative alternative. Treatments A, B, and C were otherwise the

same except that no cooperative mode of behavior was available in A, and the value of mutually cooperative behavior was higher in C than B. The data (Table 1) indicate that the median ratio of K:I was largest in Treatment A and smallest in C. As expected, the choice to cooperate appears to inhibit the choice to compete more than it does the choice to engage in individualistic behavior. The authors, however, note that these findings are not statistically significant due to the large variance.

Effect of economic value. In order to examine the effect of different values for different actions, four conditions, all with voluntary-symmetric inspection, were ordered in terms of the relative value of non-K to K. Specifically, these conditions were expected to be ordered from competitive to less competitive in the following manner: Condition B > Condition C > Condition D > Condition E. The relevant data are in Table 1. Analysis of the CNC indexes of pairs of subjects in these four treatments supports this hypothesis ($L = 268$, $p < .05$). The number of attacks also tended to increase as the value of

* For explanation of this test see Page (1963).

TABLE 1
SUMMARY OF DATA FROM NINE EXPERIMENTAL CONDITIONS

Condition ^a	Mean percentage of production			Median no. attacks for pairs ^b	Median no. inspections for pairs	Mean CNC index for pairs	Mean joint profit for pairs	Median ratio (K/I) for pairs in conditions A, B, C
	I	C	K					
A (I = 3, K = 9, VS)	.198		.802	2.5	.5	.601	99.62 ^d	7.22
B (I = 3, C = 6, K = 9, VS)	.164	.276	.560	2.8	1.5	.164	92.42 ^d	5.35
C (I = 3, C = 9, K = 9, VS)	.212	.302	.486	2.5	1.8	-.039	98.08 ^d	2.38
D (I = 6, C = 6, K = 6, VS)	.445	.201	.354	.8	1.0	-.214	106.55 ^d	
E (I = 6, C = 9, K = 6, VS)	.287	.448	.265	2.0	2.5	-.365	110.31 ^d	
F (I = 6, C = 9, K = 6, NI)	.310	.252	.438	1.3	(0)	-.173	(3.51) ^e	
G (I = 6, C = 9, K = 6, VA)	.340	.392	.268	.5	1.0	-.499	(3.48) ^e	
(Ss with inspection)	(.464)	(.404)	(.132)	(0)		(-.458)	(4.89) ^e	
(Ss without inspection)	(.266)	(.465)	(.269)	(0)		(-.566)	(2.55) ^e	
H (I = 6, C = 9, K = 6, Mand)	.282	.378	.340	1.5	(3) ^e	-.362	(2.35) ^e	
I (I = 3, C = 6, K = 9, Mand)	.086	.370	.534	2.5	(3) ^e	.123	(5.54) ^e	

Note.—The reader should not attempt to transform the mean percentage of production scores into the CNC index or the K/I ratio since the average of the ratios does not equal the ratio of the averages.

^a VS = Voluntary-symmetric inspection; NI = No inspection; VA = Voluntary-asymmetric inspection; Mand = Mandatory inspection.

^b N = 10 pairs in all conditions except F, where N = 11.

^c These were all mandatory.

^d Profits for individuals were transformed into standard scores using the following formula: $50 + \left(\frac{z - \bar{z}}{s} \right) 10$.

^e In all the above conditions the payoffs were the same (I = 6, C = 9, K = 6); therefore, standard scores were not used in the analyses involving only these conditions.

K increased but the differences are not significant. In contrast, the number of requests for inspection ($L = 266$, $p < .05$) as well as the joint profit ($L = 284$, $p < .05$) decreased as the relative value of K increased. The tendency for the number of requests for inspection to be ordered in the predicted fashion was strong enough to be statistically significant in spite of the out of sequence median in Condition D.

Availability of inspection procedures. Comparisons of the four different conditions of inspection (no inspection, voluntary-symmetric inspection, voluntary-asymmetric inspection, and mandatory inspection) were conducted under the same payoff structure. These data appear in Table 1. Four separate comparisons were made: voluntary-symmetric with mandatory inspection and no inspection with each of the three others. The results indicate that mandatory inspection yielded more joint profit than either voluntary-symmetric ($t = 2.140$, $p < .05$) or no inspection ($t = 2.600$, $p < .05$). In comparison to no inspection, voluntary-asymmetric inspection also yielded more profit ($t = 1.945$, $p < .05$) as well as a lower CNC index ($t = -1.910$, $p < .05$). However, none of the other comparisons produced significant differences in terms of either the number of attacks or the CNC index. Of particular note is the lack of any significant differences between voluntary-symmetric inspection and no inspection.

A comparison of Conditions B ($I = 3$, $C = 6$, $K = 9$, voluntary-symmetric) and I ($I = 3$, $C = 6$, $K = 9$, mandatory) further demonstrates that mandatory inspection yielded higher joint profits than voluntary-symmetric inspection ($.10 > p > .05$). Once again, however, there are no differences between these conditions when one compares the number of attacks or the CNC indexes.

Trends in Behavior

Attacks. While the preceding analyses provide information regarding which conditions lead to more or less competitive behavior, they do not permit one to examine a number of interesting dynamic aspects of the interaction. The production of weapons, for example, was influenced not only by the experimental variables but also by events which

occurred during the interaction. Two are of particular interest: attacks and inspections. Table 2, which is based on data in all voluntary-symmetric inspection conditions (i.e., Conditions A, B, C, D, and F), contains the mean percentage of weapons produced before and after the first attack. These data, presented for both attacker and attacked, are based on conditions where the first interaction was an attack rather than an inspection. (The postattack occurred after the next attack, or until there was an inspection, or until end of the game—whichever came first.)

These data indicate that the proportion of weapons produced prior to the first attack was significantly less than the proportion produced after the attack ($F_{1,84} = 4.76$, $p < .05$). Attackers produced a larger percentage of weapons than did those attacked ($F_{1,84} = 5.610$, $p < .05$). After an attack, however, the difference between them was considerably reduced. In contrast, Table 3 shows that there were no significant differences in the proportion of weapons produced by the inspector or the person inspected. (Table 3 presents data based on pairs in which the first interaction was an inspection rather than an attack.) Similarly, there were no differences in the proportion of weapons produced in the preinspection and postinspection periods. Attacks increased weapon production significantly: inspections did not.

A more detailed analysis of these data is obtained by examining the difference in the proportion of weapons produced before and after an attack by attack winners and losers. Here, we find that attack losers (mean percentage of increase = .043) increased the proportion of weapons which they produced

TABLE 2

MEAN PERCENTAGE OF K BY ATTACKERS AND BY THOSE ATTACKED IN PREATTACK AND POSTATTACK PERIODS*

Attack period	Mean percentage of K	
	By attacker	By person attacked
Pre	.595	.388
Post	.671	.582

* There were 22 subjects in each cell or a total of 88.

TABLE 3

MEAN PERCENTAGE OF K BY INSPECTIONS AND BY THOSE INSPECTED IN PREINSPECTION AND POSTINSPECTION PERIODS*

	Mean percentage of K	
	by inspection	by person inspected
Pre	215	197
Post	242	201

* There were 25 subjects in each cell for a total of 100

significantly more than did attack winners (mean percentage of increase = .295; $t = 2.709$, $p < .01$, two-tailed test). In part, these results reflect the fact that the attack losers, who produce fewer weapons to begin with, are usually the attacked rather than the attackers.

Inspections. Let us briefly turn our attention to the effect of inspections. If one only considers the number of weapons that each party has, there are three conditions that can occur when one person requests information about the other. He can find that he has more, less, or the same amount of weapons as the other. Surprisingly, the proportion of weapons produced is not significantly affected by whether the subject learns that he is producing more, less, or the same amount as the other player.

It is clear, however, that attacks generally lead to further increases in the production of competitive products while inspections do not. Since we have assumed that there is a mutually reinforcing relationship between the production of weapons and attacks, we should expect to find differences in the kinds of future actions that attacks and inspections precipitate. Table 4 presents the relevant data. If the first action is an attack it is evident that the action which is most likely to follow it is an attack. Whereas, if the first action is an inspection, the next action will probably be an inspection.

Profits. Profit is one of the primary indicators of success in games of this variety. The data have already indicated that profit tends to be associated with the relative value of K and non-K, and the kind of inspection system which is operating. These data have also indicated that there is an inverse rela-

TABLE 4

CHI-SQUARE ANALYSIS OF KINDS OF ACTIONS PRECIPITATED BY ATTACKS AND INSPECTIONS*

First action	1st action	
	Attack	Inspection
Attack	4	2
Inspection or end of the game	1	1

$$\chi^2 = 6.611^{**}$$

* Based on Conditions A, B, C, D, and E
** $p < .01$

tionship between competitive behavior and profit. The authors studied the impact on profits by examining the relationship between attacks and profit.

This analysis was conducted in two ways. First, the relative number of attacks and the relative profit within pairs of subjects were compared. Then the relationship between the number of attacks and profits between pairs of subjects was examined. The data indicate that those who attacked more frequently than their opponents were more likely to make greater profits than their less aggressive counterparts ($Z = 2.78$, $p < .01$, two-tailed binomial test). The data, however, show that this interpretation cannot be generalized to pairs of subjects. The profits of players, who were as aggressive, or more aggressive than the person with whom they were paired in more aggressive dyads (mean profit = 50.157), were not significantly different than the profits of their counterparts

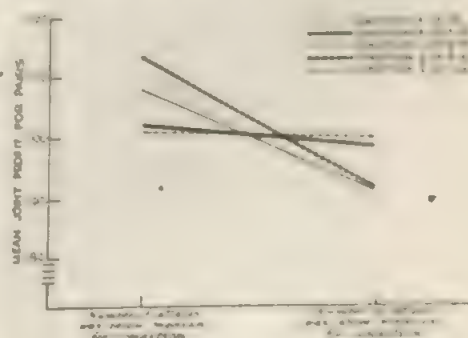


FIG. 1. Relationship between number of attacks and joint profit of pair. (Profit was computed for individuals in terms of standard scores: $50 + (x - \bar{x})/s$ 10.)

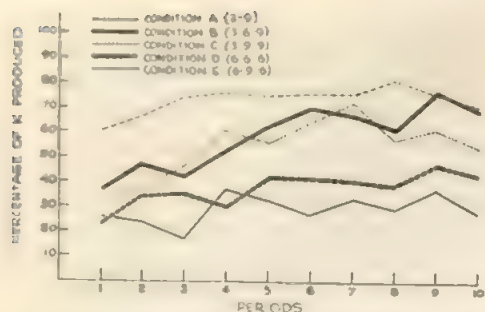


FIG. 2. Percentage of K over 10 trial periods for various payoff conditions with voluntary symmetric inspection.

in the less aggressive dyads (mean profit = 53.175). More strikingly, the data in Figure 1 suggest that aggression may be even more harmful under conditions with relatively low competitive incentives than it is under high ones. Subjects who are highly competitive under conditions not normally eliciting much competition do less well than subjects who are highly competitive under conditions which stimulate competition.

Competitive spiral. The data cited above suggest that competitive incentives lead to competitive behavior and aggression, and that aggression may lead not only to more aggression but also to less profit. With this competitive spiral in mind let us turn to the trends in the relative percentages of K over the 10 2-minute periods. Figure 2 contains these data.

As one would expect, K tends to be greatest when there is no opportunity to choose C, and steadily decreases as the relative value of K decreases. Analysis of the slopes of the data presented in Figure 2 suggests that the stability of the interaction was different in the different conditions ($F_{4,45} = 3.02, .01 < p < .05$).⁴ The percentages of weapons produced in Condition A, where no cooperative alternative was available, and in the most cooperative of the conditions, Condition E, were comparatively stable over the 10 periods. The slopes of these curves suggest that the stability of the interaction may be related to the clarity of the structure. This interpretation is supported by a comparison of the two most

⁴ For discussion of this analysis see McNemar (1962).

stable conditions (A and E) with the three least stable ones (B, C, and D). In the former, the economic structure clearly favors either K or non-K. In the latter, the economic structure, in varying degrees, is less clearly one or the other. The authors suggest that when incentives are clearly cooperative or competitive, each of the participants expects the other to react to the particular incentives. In both instances, their expectations are likely to be confirmed, and, therefore, little change in behavior will result. In other conditions (particularly in Condition B, the least stable condition), the incentives are ambiguous. As a result, subjects have unclear and not very strong expectations of the other. The onset of competitive behavior, however, provides evidence of the other's intention and elicits a competitive response.

This interpretation of the data suggests that the initial concordance of behavior within pairs of subjects will differ in the various experimental treatments. Namely, the differences between subjects in the percentage of K should be least in Conditions A and E (the most stable ones), and the difference should be greatest in Conditions B, C, and D (the least stable conditions). Comparing these two sets of means, based on activity in the first (.20 > p > .10) and last (.10 > p > .05) 8 minutes of play, provides tentative support for this explanation of these findings (Table 5).

Conditions A and E not only had the highest concordance of behavior early in the game, but there was some tendency for them to maintain this position late in the game.

TABLE 5
MEAN DIFFERENCE IN PERCENTAGE OF K DURING
FIRST AND LAST EIGHT MINUTES
OF ALLOCATION GAME

Condition	Mean difference in percentage of K within pairs		No. pairs
	1st 8 minutes	Last 8 minutes	
A (I = 3, K = 9)	16.2	24.2	10
B (I = 3, C = 6, K = 9)	25.9	30.6	10
C (I = 3, C = 9, K = 9)	22.6	26.0	10
D (I = 6, C = 6, K = 6)	23.3	24.6	10
E (I = 6, C = 9, K = 6)	15.5	7.8	10

(This was somewhat less true for Condition A than for Condition E.) Similarly, Condition B had the lowest concordance of behavior both early and late in the game. Thus, it seems reasonable to suggest that Conditions A and E initially lead to more mutuality of expectation and choice. Subsequent affirmation of these early expectations and actions has a stabilizing effect on the interaction.

Discussion

Effect of a cooperative alternative. It is not surprising to find that providing the subjects with a cooperative alternative increases the percentage of noncompetitive behavior. These data, however, not only support this self-evident notion, but they also suggest that doing so tends to reduce the relative amount of their competitive behavior. This point is critical since it is possible that, after introducing a cooperative alternative, subsequent increases in cooperative interaction may occur without a decrease in competitive interaction. For example, let us consider two siblings who relate in two ways, competitively or not at all. If at one point their parents suggest a mode of cooperative behavior, they may accept the advice. However, we cannot conclude that they will engage in less competitive behavior. Instead, they might spend less of their time apart, more of their time in cooperative play, and the same amount of time in competitive interaction. By analogy, the data which deal with the K/I ratios suggest that this is unlikely. The addition of a cooperative alternative is more likely to lead to proportionately more noncompetitive and less competitive behavior. This interpretation of the data is supported further inasmuch as the mean percentage of K ($\bar{X} = .806$) in Condition A (no cooperative alternative) is significantly greater than the mean percentage in Condition B ($\bar{X} = .560$, $t = 2.42$, $p < .05$), even though the average number of products in Condition A ($\bar{X} = 50.8$) is greater than in B ($\bar{X} = 43.2$).

Effect of economic incentives. The data provide strong support for the assumption that the allocation of efforts among noncompetitive and competitive alternatives is af-

fectured by the values which are immediately associated with the alternatives. However, it is evident that subjects chose not only in terms of what outcomes were more desirable but also in terms of what they expected the other subject to do. When the payoff structure does not lead to clear initial expectations of competition or cooperation, the interactional process causes the ambiguity to be resolved increasingly in terms of competition. This finding is in keeping with a good deal of prior research which supports the proposition that a state of trust or cooperation is easier to transform into a state of suspicion or competition than vice versa (Deutsch, 1958; Deutsch & Krauss, 1960; Solomon, 1960).

Effects of inspection. Providing players with the opportunity to inspect each other, without requiring inspection—as in voluntary-symmetric inspection conditions—had no noticeable advantages over the no-inspection situation. Mandatory inspection, however, consistently produced higher joint profits than either the no-inspection or voluntary inspection conditions. These findings suggest the possibility that, as the perceived likelihood of being inspected increases, the readiness to engage in aggressive behavior may decrease. This interpretation receives some additional support if one is willing to assume that the subject who could not inspect the other in the voluntary-asymmetric condition perceived that the likelihood of being inspected was relatively high. In this condition one finds that not only are joint profits higher than in the no-inspection condition, but the subject without inspection potential had the lowest CNC index of all subjects and also produced the highest percentage of cooperative products. A possible explanation is that as he perceived that likelihood of inspection would increase, he attempted to induce the other to refrain from being competitive or aggressive by demonstrating a lack of competitive intent.

Trends in competitive behavior. The findings of this experiment have supported the assumption that the relative economic value of competitive and noncompetitive behavior will influence not only the subjects' preferences but also their mutual expectations and

behavior toward one another. When there are strong incentives favoring either competitive or noncompetitive behavior, the players have clear expectations of one another, and consequently their behavior is relatively more concordant. When the incentives do not clearly favor either strategy, there is more likelihood that cooperation or competition will not be mutual. The lack of mutuality is unstable and tends to turn into mutual competition rather than mutual cooperation. Consequently, only the relatively more concordant-cooperative conditions (Condition D and Condition E) lead to a sustained cooperative interaction.

The joint profits of subjects in the more competitive conditions are lower than their joint profits in the more cooperative conditions. The decrement in joint profits, however, is greater when aggressive behavior occurs under cooperative rather than competitive conditions; that is to say, when opponents initially have highly competitive orientations, subsequent competitive behavior, such as attacks, has little effect on outcomes. Such action only increases the degree of competition. However, if aggression occurs when the participants are more cooperatively

oriented, the action is a change in kind; it is incongruous with the cooperative context within which the interaction is taking place. It seems reasonable to suggest that expectations and trust have been violated when this occurs and that the resulting response is highly competitive. Consequently, we find the joint profits of the more aggressive pairs in cooperative conditions to be even lower than the joint profits of the more aggressive pairs in competitive conditions.

REFERENCES

- DEUTSCH, M. Trust and suspicion. *Journal of Conflict Resolution*, 1958, 2, 265-279.
- DEUTSCH, M., & KRAUSS, R. M. The effect of threat upon interpersonal bargaining. *Journal of Abnormal and Social Psychology*, 1960, 61, 181-189.
- LUCE, R. C., & RAIFFA, H. *Games and decisions*. New York: Wiley, 1958.
- MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1962.
- PAGE, E. B. Ordered hypotheses for multiple treatments: A significance test for linear ranks. *Journal of the American Statistical Association*, 1963, 58, 216-230.
- SOLOMON, L. The influence of some types of power relationships and game strategies upon the development of interpersonal trust. *Journal of Abnormal and Social Psychology*, 1960, 61, 223-230.

(Received May 25, 1966)

REACTIONS TO SUFFERING OF OTHERS UNDER CONDITIONS OF INDIRECT RESPONSIBILITY¹

MELVIN J. LERNER

AND

GAIL MATTHEWS

*Department of Behavioral Science
University of Kentucky*

University of Kentucky

The experiment was presented as a study of paired-associate learning to 66 female undergraduates who participated in pairs. Each S, with the exception of those in a control condition, was led to believe that her partner was to receive a strong electrical shock and that S would be in a more desirable condition (control or money). When S perceived that the other person was responsible for her own suffering, subsequent descriptions of the other person were relatively objective. When S perceived herself as responsible for the other person's fate, she tended to devalue her. When the other person was perceived as responsible both for her own suffering and for S's desirable fate, the attractiveness of the other person was enhanced.

This study is the second in a series concerned with examining those factors which determine whether a suffering victim is reacted to with indifference or rejection rather than concern and affection. The first study (Lerner & Simmons, 1966) found that when observers were unable to alter the fate of an innocent victim so that the suffering ended and the victim was compensated, they rejected the victim by describing her as an undesirable person. A number of earlier studies (Davis & Jones, 1960; Glass, 1964; Lerner, 1965a) have established that if a person has actually inflicted suffering on someone, he is likely to reject and devalue the victim. Apparently the persecutor justifies his actions by persuading himself that the victim deserved his fate.

Although these studies provide some understanding of the factors which influence the person's reaction to a victim when the person is either a "persecutor" or "observer," there is another kind of relationship which merits examination. Whether actually true or not, many people perceive themselves tied to others in a kind of contest or competition. The underlying assumption is that there are a

limited number of desired resources available, and someone has to lose out if others are to succeed. From this perspective it is possible to consider the fate of deprived and suffering people as the result of the choice of those more privileged to maintain what relative advantages they have. The privileged resist serious reductions in their status which would better the fate of those suffering. This implies a kind of responsibility on the part of the privileged people for others' suffering. But certainly, this level of responsibility does not reflect any direct acts designed to cause the suffering.

The present study examines the consequences of this rather indirect level of responsibility for the suffering of another person. To approximate the desired conditions in a laboratory setting, pairs of students who had agreed to participate in a study of human learning were faced with the prospect of one of them having to be in a condition of negative reinforcement (strong electric shock) and the other in a control condition in which they merely received appropriate feedback about their answers. The decision as to which of them would be in the negative reinforcement and which in the control was to be determined solely by which of two slips of paper each subject selected out of a bowl. The subjects were led to believe one slip always contained the word "shock" and the other "control." It was apparent, then, that the fate of one subject was clearly tied to the fate of the other

¹This research was supported by a grant from the National Science Foundation (GS-957) administered by the senior author.

The authors wish to express their great appreciation to Donna Crockett and Rosemary Lichtman for their assistance in gathering and analyzing the data and to John Haley, who provided valuable advice concerning the statistical analyses of the data.

so that the subject who picked first determined not only her own fate, but the other person's fate as well. When the subject picked first (and selected control, by chance) she was, in a sense, responsible for the other receiving electric shock. When the subject was led to believe the other picked first and had selected negative reinforcement, again by chance, the subject was able to believe that the other was responsible for her own suffering and the subject's good fortune.

The important hypotheses in the study are concerned with the subject's reaction to the other person in these conditions. A set of predictions was derived from the theoretical notions developed earlier by Lerner (Lerner, 1965b; Lerner & Simmons, 1966). The underlying process postulated in these papers is based on the person's need to believe in a just world in which people deserve their fate. It was also reasoned that the person would prefer to believe that people merit their fate because of what they have done. It is only when the person cannot ascribe some misdeed to a suffering victim that he will attempt to maintain his belief in a just world by persuading himself that the victim deserves his suffering by virtue of his being an undesirable, bad person.

It follows from these notions that the condition in which the other person determined her own fate (by picking first) should elicit little, if any, need to reject the victim. The victim's suffering can be attributed to her own behavior. On the other hand, when the subject picks first, she should perceive herself to be responsible for the victim's suffering. The subject then will be compelled to justify the state of affairs by attributing undesirable characteristics to the other.

To provide additional information, some subjects were presented with the alternatives of negative reinforcement and positive reinforcement (\$.25 for each correct answer) rather than negative reinforcement and a control condition. Again, all subjects were led to believe they would be in the desirable condition. If, in addition to causing the other to suffer, the subject benefits herself at the expense of the other, then this might elicit a different reaction to the victim than when the subject merely avoids the suffering herself

(shock/control). As a result of a contrast effect, the subject might perceive the victim suffering as greater (Heider, 1958, pp. 281-286) and therefore exhibit greater rejection of the victim. On the other hand, the subject resting securely in the midst of her good fortune, could afford to feel more compassionate toward the suffering victim. Or, the subject's good fortune in obtaining money might be superfluous to her good fortune in avoiding the suffering which faces the victim. If this is the case, then the addition of money should make no difference in the subject's reaction to the victim.

To assess the effect of this indirect level of responsibility on attraction to a victim, it is necessary to provide some conditions in which the other person causes her own suffering, yet does not at the same time benefit the subject. This was done by having some subjects choose a slip from a larger bowl full of slips—half of which indicated negative reinforcement and half the more desirable condition (either control or positive reinforcement). Although all subjects were led to believe they were in the desirable condition and the other in the negative reinforcement, their respective fates were arrived at independently (fates independent).

One further condition was included in which the subjects were led to believe that both of them would be run under control conditions (control/control). This control group was used to test the effects of the fates-independent condition. In the fates-independent condition, rejection of the victim may occur merely because the subject learns that the other will suffer. On the other hand, if the perception that the other subject is responsible for her suffering eliminates the need to reject her, then there should be no difference between the subject's rating of the other's attractiveness in this control/control condition and in the fates-independent condition.

Unlike the previous studies cited, the present one also explored the relation between the subject's personal evaluation of the victim and the subject's willingness to attempt to reduce the victim's suffering. This was tested by allowing the subject to believe that the victim was extremely frightened and had asked for someone to keep her company while she awaited the onset of her being shocked.

The subjects were given an opportunity to comply with this request after having made their ratings of the victim. The most obvious prediction was that the probability of a person helping a victim would be related to her attraction to the victim. The more attracted to the victim the more likely one is to render help.

The conditions described above can be summarized into a 2×3 factorial design with the addition of one extra control condition. There were two main conditions of outcome: (a) negative reinforcement versus control (shock/control), (b) negative reinforcement versus positive reinforcement (shock/money); and three conditions of determining the outcome: (a) the subject determined the outcomes (self picks first), (b) the other determined the outcomes (other picks first), and (c) the outcomes were determined independently (fates independent). The additional condition was one in which both subjects were in the control condition (control/control).

The effects of the fates-interdependent conditions (other picks first and self picks first) will be tested against the fates-independent condition if there are no appreciable differences between the attractiveness ratings of the other subject in the fates-independent and control/control conditions. The main hypotheses were:

1. The other person will be rated as more attractive in the other-picks-first condition than in the self-picks-first condition.
2. The other person will be rated as less attractive in the self-picks-first condition than in the fates-independent condition.
3. Those conditions which provide the most positive ratings of the other person will provide the most subjects who volunteer to help the other person.

In addition, the differences between the shock/money and shock/control conditions will be examined.

METHOD

Subjects

The subjects were 66 female students who volunteered to participate in an experiment on human learning as part of the expected requirements for a course in introductory psychology. They were exposed to the experimental situation in pairs, pre-

assigned on a non-systematic basis to one of the various conditions. In those cases where one subject did not appear at the assigned time, a confederate of the experimenter took the place of the second subject. The data from eight subjects were not included in the final sample. Five of these doubted some part of the experimental ruse, one was excluded because she had learned of the nature of the experiment, and two were eliminated because they had extended contact with one another prior to the experiment.

Procedure

After the subjects had gathered in a waiting room, the experimenter ushered them into one of two experimental labs. There the subjects were told that they were to take part in a study concerning the effect of negative reinforcement on paired-associate learning. As the experimenter explained the purpose of the experiment she pointed to a massive drum, an imposing apparatus for delivering strong electric shocks through wrist electrodes, and an impressive-looking oscilloscope.

The subjects were then put in separate waiting rooms in anticipation of being run in separate labs. Once separated, the subjects learned what conditions they were to undergo—negative reinforcement or some other more desirable condition—and completed some scales which were designed to allow the subject to describe the attractiveness of the person she had just met.

While these measures were being obtained the subject also responded to questions concerning her reaction to the experiment up to that point. A few moments later while the experimenter was explaining the next part of the experiment to the subject, they were interrupted by a second experimenter who summoned the experimenter into the hall. The second experimenter confirmed the fact that the experimenter was supposed to vacate that particular room shortly so that another experiment could take place there. In addition, the other experimenter informed the experimenter that she had picked up the other subject's scales for the experimenter and had noticed what that subject had written in response to the "reaction to the experiment" questions. The second experimenter then read the experimenter the other subject's comments to the effect that she was terrified of being alone while awaiting the shocks and she desperately wanted someone to talk with and keep her company. The experimenter merely nodded affirmation and agreed to leave the lab immediately. Although it was arranged so that the subject would be compelled to overhear this bit of interaction, it was intended that the subject not believe the discussion was directed, in any sense, to her. Shortly after this, the experimenter then told the subject she must wait somewhere else, and as they left, asked the subject where she preferred to wait—in another room with some magazines and books or in the room she had been in initially with the other subject.

After the subject made her choice, she was interviewed about the bases of her choice and about what she had been thinking throughout the experiment. She was then informed of the true nature of the experiment.

Responsibility for the Outcomes: Fates Interdependent

While receiving the initial orientation the subjects were told that one subject in each pair would be in the negative-reinforcement condition and the other subject in a control condition (or positive reinforcement). The subjects were told that the condition each would be in depended upon which of the two slips of paper they selected out of a bowl. They were told one slip contained the word "shock" and the other "control" (or positive reinforcement). It was clear to all subjects that the subject who selected the first slip would determine the fate of both. Soon after the subjects were separated, the experimenter entered the subject's room with the bowl containing the slips.

In the self-picks-first condition, the experimenter entered with two slips in the bowl and informed the subject that she would pick first. Since the bowl contained only slips with "control" (or positive reinforcement) written on them, the subject was led to believe she was in that condition and therefore the other subject in the shock condition.

In the other-picks-first condition, the experimenter entered the room with the bowl containing one slip. The subject was told that the other subject was allowed to pick first. The subject then took the remaining slip and discovered her good fortune.

Responsibility for the Outcomes: Fates Independent

This condition was essentially similar to the fates-interdependent conditions except that the subjects were presented with a large bowl containing numerous slips of paper and told that half the slips designated the negative-reinforcement conditions and half the control or positive-reinforcement condition. When the subject informed the experimenter what condition she had picked, the experimenter remarked casually that the other subject had picked one of the slips indicating shock.

Degree of Differential Outcomes

The anticipation of differential outcomes was created by informing the subjects that the alternatives were either shock versus control or shock versus money (\$.25 for each correct answer). In one condition the subjects did not pick slips out of a bowl but were merely informed that they both would be run in control conditions.

Measures

After entering their separate rooms and learning of their fate, the subjects completed two scales designed to assess their comfort and anxiety at that

point in the experiment. These were followed by scales designed to assess the attractiveness of the other subject and the subject's perception of responsibility.

Two different measures of the attractiveness of the other subject were employed:

1. Ratings on 15 highly evaluative bipolar scales (e.g., likable-unlikable, mature-immature). The ratings on these scales were combined to yield an overall index of attractiveness. The range of possible scores was from 15 to 135 (the higher the score the more positive the rating). The attractiveness rating the subject ascribed to "the average college student" was subtracted from that ascribed to the other subject to yield the final measure used in the analysis.

2. The subjects rated the other subject in response to five questions about her "Social Stimulus Value" (see Lerner & Simmons, 1966, for a more complete description of this scale).

To assess the perception of responsibility, the subjects filled out scales which allowed them to assign varying degrees of responsibility to the experimenter, the other subject, and themselves for each of their fates. They then were required to designate one of the three (experimenter, other subject, or self) as primarily responsible for the subject's fate and then for the fate of the other person. This designation of primary responsibility served as the data to determine perceived responsibility.

The subject's desire to help the suffering other was assessed by coding her answers to the experimenter's query as to where the subject wished to await the onset of her experimental condition into three categories: (a) with the other subject, (b) with books and magazines, (c) no preference. These last responses consisted of the subject's refusing to make a choice after the experimenter asked the question a number of times. Each time the experimenter refused to decide for the subject.

RESULTS

Responsibility

Although not specifically stated as a hypothesis, underlying all the other predictions was the expectation that the procedure of having one subject pick first would create an impression of responsibility for the fate of both subjects. The data in Table 1 indicate that this manipulation did, in fact, create the impression of responsibility. In the interdependent conditions, when the subject picked first out of the bowl, she tended to perceive herself as primarily responsible for both her own

² Although this scale was very effective in the Lerner and Simmons study, for some reason, as yet unknown, a number of subjects in the present study resisted having to respond to the items in this scale. As a consequence, the responses to these scales were not included in the analyses of the data.

TABLE 1

NUMBER OF SUBJECTS ATTRIBUTING PRIMARY RESPONSIBILITY TO EXPERIMENTER (E), SELF (S), AND OTHER (O)

		Fates independent	S picks 1st	O picks 1st
Responsible for S's fate ^a	E	3	2	1
	O	0	0	12
Responsible for O's fate ^b	S	12	13	5
	E	5	2	2
	O	9	6	15
	S	1	7	1

^a Comparing self picks first with other picks first, while eliminating E choices, yields $\chi^2 = 12.50$, $df = 1$, $p < .001$.

^b Comparing self picks first with other picks first, while eliminating E choices, yields $\chi^2 = 5.91$, $df = 1$, $p < .02$.

fate and the other subject's fate as well. When the subject was led to believe the other person picked first, the other was assigned primary responsibility for the outcomes. In the fates-independent conditions, the subjects perceived themselves as responsible for their fate and the other person as responsible for hers.

Attractiveness of the Other Person

Since the mean ratings of the other subject in the fates-independent condition ($\bar{X} = -4.41$) and the control/control condition ($\bar{X} = -6.30$) are quite comparable, Hypotheses 1 and 2 will be tested employing the fates-independent condition as the more appropriate control (See Table 2).

Hypothesis 1 was supported by the data. The other subject was described as considerably more attractive in the other-picks-first condition than in the self-picks-first condition

TABLE 2

\bar{X} RATINGS OF THE OTHER PERSON'S ATTRACTIVENESS: BIPOLAR ADJECTIVES^a

Situation	Shock/control	Shock/money	Control/control
Fates independent	-5.25 (<i>N</i> = 8)	-3.57 (<i>N</i> = 7)	-6.30 (<i>N</i> = 10)
Other picks 1st	6.55 (<i>N</i> = 11)	3.14 (<i>N</i> = 7)	
Self picks 1st	-11.22 (<i>N</i> = 9)	-19.33 (<i>N</i> = 6)	

Note.—Variance estimates were $MS = 770.53$, $F(6) = 4.36$, $p < .005$ (between cells), and $df = 51$, $MS = 176.90$ (within cells).

^a Other subject—average college student.

($t = 4.32$, $df = 51$, $p < .001$). The other person was also described as more attractive when she picked first, determining both fates, than when the outcomes were determined independently (other picks first versus fates independent, $t = 2.12$, $df = 51$, $p < .05$). Hypothesis 2 was also confirmed. When the subject was led to believe she was responsible for the other person receiving electric shocks, she rated the other as less attractive than when the other determined her own fate (self picks first versus fates independent, $t = 2.05$, $df = 51$, $p < .05$).

Although the mean ratings of attractiveness in the shock/money conditions tend to be lower than in the shock control conditions, the differences are not significant ($t = 1.22$, $df = 51$). Apparently the addition of the certainty of earning some money had no appreciable affect on the subjects rating the victim's attractiveness.

Decision to Help

The great majority of subjects in all conditions in which the fates were interdependent decided to return to the room with the other person who wanted someone to comfort her (see Table 3). The prediction from Hypothesis 3 would have been that more subjects in the other-picks-first condition than in the self-picks-first conditions would have made this choice since the victim was rated more attractive in the other-picks-first condition. These results become more interesting, however, when compared with the choices in the fates-independent conditions. Fewer subjects in the fates-independent conditions chose to return to the suffering other than in the conditions where the fate of the subject had been tied to that of the other person ($\chi^2 = 6.09$,

TABLE 3

NUMBER OF SUBJECTS SELECTING ALTERNATIVE PLACES TO AWAIT ONSET OF EXPERIMENT

Choice of place	Experimental conditions		
	Fates independent	Other picks 1st	Self picks 1st
With other	4	12	11
With books	1	2	1
No preference	10	4	3

$dt = 1, p < .02$). This relationship held even though the other person was described as more attractive in the fates-independent than in the self-picks-first conditions.

DISCUSSION

The findings of this study clearly indicate that the perception of responsibility for one's own and another's fate can be elicited in an extremely indirect fashion: by the chance selection of one of two possible slips of paper which indicate either a desirable outcome or one involving pain and suffering. The evidence for this conclusion comes not only from the subject's direct report of responsibility, but even more importantly from the effects of the different conditions of responsibility on the subject's reaction to the other person.

Given that the perception of responsibility was elicited by the indirect technique, some of the results were not surprising. When the subject believed she was responsible for the other person's suffering she tended to devalue the personal characteristics of that person. This merely confirmed the findings of previous studies which employed more direct means of creating the cognition that a person has harmed someone.

Of more importance were the findings relating to the perception of the other person as responsible for her own suffering. The findings support the assertion that rejection of a suffering person will not occur if the observer can attribute the suffering to something the victim did. When the other person selected the slip which designated the shock condition and her choice had no effect on the subject's fate (fates-independent condition), the subject described the other person as no less attractive than in the control/control condition where neither of them was to be shocked. Apparently the other person was perceived as responsible for her own suffering, and the subject's need to believe in a just world was not threatened.

This same reasoning led to the prediction that when the other picked first, she would not be rejected. It was implied, though not stated, that the other person would be perceived as equally attractive in the other-picks-first and in the fates-independent conditions. The subject's perception of the other would

be fairly objective in both of these conditions. However, this did not occur. Surprisingly, the subjects enhanced the attractiveness of the other when that person was perceived as responsible for her own suffering and the subject's good fortune. What seems to have happened is that the subjects became grateful to the other person in this condition. It may be that this condition elicited the kind of intimacy and gratitude that is associated with the idea of, "There, but for the grace of God go I."

If this last conjecture is true, then the commonly held notion that identification with a victim is an important determinant of whether the person will respond with compassion or rejection begins to take on some meaning. Previously, Lerner and Simmons found no relationship between the perception of similarity with a victim in terms of personal attributes and attitudes, and the person's rejection or acceptance of the victim. The present study seems to indicate that identification with a victim requires the perception of the same possible common fate and not the perception of similar attributes.

An additional surprising and equally interesting finding was the large number of subjects who chose to return to the suffering victim in both conditions in which their fates had been interdependent. This held even among those who tended to perceive themselves as responsible for the suffering and who had clearly devalued the character of the victim.

One possible explanation for this finding is that the choice to return to the suffering victim did not reflect, in all subjects, a desire to help the victim in response to her plea. Rather, some subjects may have been merely curious and others may have been hoping to derive some sadistic pleasure from the sight of the victim's suffering. Although almost all subjects, when interviewed, gave the desire to comfort the victim as the basis of their choice to return, this does not necessarily negate the possible presence of these other motives.

The subjects were asked in their postexperimental interview why they made the choice they did. Although no definitive answers could be found in their responses, there did seem to be some differences between the

growers given by those subjects in the self-pick-first condition and those in the other-pick-first. Those subjects in the self-pick-first who felt somewhat responsible for the other's suffering, tended to indicate curiosity or a desire to assuage the other's hostility as reasons for going back to her (e.g., "... to see whether what I said about her was true, she probably would have said 'you lucky rat!'"). Those subjects in the other-pick-first condition who had enhanced the other's attributes seemed to show a kind of identification with the other (e.g., "... because I heard she was upset and I was completely calm—thought I could help her. I know if I had felt the same way I would have wanted someone with me.").

Obviously the presence of these and other possible motives indicates the importance of exploring, in a systematic fashion, the relationship between a person's attitude toward a victim and those factors which determine the likelihood of administering aid and comfort, even at the expense of one's own well-being. From one perspective it matters little if the victim's suffering is alleviated by some-

one acting out of noblest obligation or some strong feeling of personal affection and concern. There may be, however, important differences in the victim's reaction to help based on differing motives.

REFERENCES

- DAVIS, K. E., & JONES, F. J. Changes in interpersonal perception as a means of reducing cognitive dissonance. *Journal of Abnormal and Social Psychology*, 1960, 61, 402-410.
- GLASS, D. C. Changes in being as a means of reducing cognitive dissonance between self-esteem and aggression. *Journal of Personality*, 1964, 32, 541-549.
- HEIDER, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.
- LENER, M. J. The effect of responsibility and choice on a partner's attractiveness following failure. *Journal of Personality*, 1965, 33, 178-187 (a).
- LENER, M. J. Evaluation of performance as a function of performer's reward and attractiveness. *Journal of Personality and Social Psychology*, 1965, 1, 155 (b).
- LENER, M. J., & SIMMONS, C. H. Observer's reaction to the "innocent victim": Compassion or rejection? *Journal of Personality and Social Psychology*, 1966, 4, 208-210.

(Received May 16, 1966)

CONTEXT EFFECTS IN COMMUNICATION

MELVIN MANIS¹

University of Michigan

Ss were given verbal descriptions of several emotional states. Some Ss attempted to match each description with the 1 photograph (out of 11) that the writer was attempting to describe; others were to rate each description on a pleasant-unpleasant continuum, using either an 11-point scale or a non-numerical equivalent. In all 3 tasks, contrast effects were produced by providing some Ss with mainly pleasant descriptions, while others responded to primarily unpleasant descriptions. These results probably reflect context-induced changes in the experience elicited by the test stimuli, rather than changes in Ss' subjective scales. Ss given an unbiased sample of descriptions were more accurate in identifying the "target" photographs than were those assigned to biased contexts.

One of the most replicable findings in the literature on subjective judgments is the fact that the individual's response to a given stimulus is partly determined by the *total array* of stimuli to which he is exposed, not solely by the stimulus being judged. This contextual effect usually takes the form of an apparent *contrast* between the overall context and the test stimulus; for example, Campbell, Hunt, and Lewis (1958) have shown that in judging the degree of disorganization and eccentricity of thought implied by various definitions on the Wechsler Vocabulary scale, subjects who have been primarily exposed to "high-pathology" definitions generally rate "midscale" definitions as indicating *less* eccentricity than do subjects whose predominant experience has been with "low-pathology" definitions. Similar effects have been reported in studies dealing with judgments of weights (Heintz, 1950), length of lines (Krantz & Campbell, 1961), and in a variety of other social and psychophysical domains (Helson, 1964).

Despite the reliability and generality of this effect, the basic data may be interpreted

¹ The author is also affiliated with the Ann Arbor Veterans Administration Hospital. All statements are those of the author and do not necessarily represent the opinions or policy of the Veterans Administration.

The author would like to thank Sally Leavy and Arnold Weingarden for their help in assembling the experimental materials. The latter also played a major role in collecting and analyzing the data. Others participating in the data collection were John Stevenson and Kenneth Wolf.

in two main ways. A *perceptual* interpretation has been offered by some researchers, based on the assumption that extreme contexts affect the *subjective impression* elicited by the various test stimuli. This would imply, for example, that in responding to vocabulary definitions, a judge whose predominant experience has been with *pathological* items will displace moderate definitions towards the *nonpathological* end of the scale; that is, these items will be "seen" as relatively well organized and lacking in eccentricity.

A rather different view, emphasizing changes in the subject's *subjective scale* (rather than changes in the experience elicited by a given stimulus), has been offered as an alternative explanation (Campbell, Lewis, & Hunt, 1958; Krantz & Campbell, 1961; Stevens, 1958; Upshaw, 1962; Volkman, 1951). According to this interpretation, an extreme context does not affect the subject's subjective experience, but instead influences the "language" which he uses to describe this experience, and thus represents a *semantic*, rather than a *perceptual*, effect. For example, consider a judge who attempts to rate the disorganization and eccentricity of various definitions that range from *extreme disorganization* to *moderate disorganization*; assume he is to use a 9-category scale with "totally disorganized" and "well organized" as end points. Since the stimuli cover a narrower range on the subjective continuum than do the available response categories, the subject may redefine the meaning of "well organized"

to make the bounds of his subjective scale more congruent with the definitions that he is shown. Such a shift would, of course, result in a repositioning of *all* the category boundaries, since the experimenter's instructions have been to divide the total subjective continuum into 9 categories of equal breadth. Figure 1 depicts these alternative explanations for context-induced contrast effects; it is important to note that both accounts are completely consistent with the experimental observation that neutral stimuli usually elicit judgments that *contrast* with the bulk of the subject's prior judgmental experience.

In discussing these two interpretations, Campbell, Lewis, and Hunt (1958) suggest that shifts in the meaning of the various response alternatives (as in Figure 1, bottom) may be particularly likely when the subject is required to make his judgments using a response language that the experimenter creates for the experiment, for the experimenter typically introduces a rather vague language that is novel, arbitrary, and relativistic. The novelty of the language implies that the subject must learn the meaning for the different response categories as he becomes familiar with the stimuli to be judged; clearly, the results of this learning process will be affected by changes in the stimulus set. The relativity of the typical response language produces similar problems. In the words of Campbell, Lewis, and Hunt (1958, pp. 220-221):

Terms like *heavy* and *light* are in their proper semantic frame, and are usually treated as such. However, to a naive observer, and from a specific point of view, a heavy suitcase is not a heavy suitcase. A heavy suitcase is not a heavy suitcase. A heavy suitcase is not a heavy suitcase. In such usage terms like *heavy* and *light* contrast with "absolute" terms like *one hundred pounds*, or *three tons*, which are dealing with the same attributes of physical objects. Such terms are extracted from specific communicative situations and are understood to be invariant attributes of the objects, appropriate descriptions of them, whatever the setting.

In brief, given the typical vague characterization for the various response categories in a rating task, there is little to prevent subjects from privately defining the various alternatives to "fit" the range of stimuli that are presented.

To demonstrate the validity of this argument, Campbell and his associates have conducted two experiments in which judgmental shifts associated with contextual differences are compared in two situations: (a) when subjects are instructed to rate the stimuli using response categories that are clearly anchored and thus relatively unsusceptible to redefinition, and (b) when the response categories are less well defined. In one study (Campbell, Hunt, & Lewis, 1957), subjects rated the amount of disorganization present in a series of vocabulary definitions. For those in the "Detailed" condition, each of the 9 points on the response scale was defined explicitly (e.g., Category 3 was to be used for definitions showing "very slight traces of disorganization and eccentricity"); for subjects in the "Simple" condition, however, only the end categories were explicitly defined. In another study (Krantz & Campbell, 1961) the subject's task was to judge the length of a series of white lines projected on a screen. Some subjects used a familiar, absolute, and presumably stable response language, making their judgments in *inches*; other subjects were instructed to use an "artificial, relativistic" language, in which the response "100" was to represent lines of "average" length, "95" was reserved for "less than average" lines, etc. In both studies, biases in the stimuli presented for judgment produced marked contrast effects when the response alternatives were unfamiliar and loosely defined; by comparison,

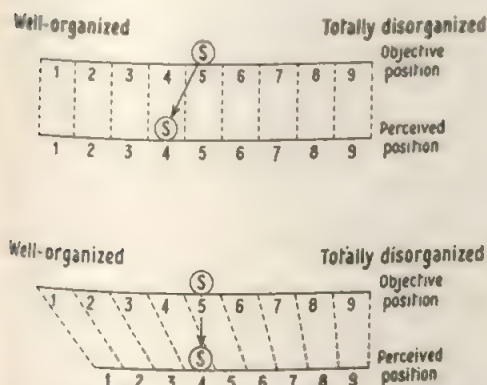


FIG. 1. Two conflicting interpretations for the phenomenon of context-induced contrast effects. (The top part of the drawing is based on an assumed perceptual effect, while the bottom depicts a semantic effect—see text.)

contextually induced contrast effects were reduced (but not eliminated) when subjects used a more firmly established set of responses. The demonstration of statistically reliable contrast effects even when subjects used a relatively unambiguous response scale argues against the claim that contrast effects are *solely* attributable to differences among subjects in their labeling of the subjective continuum; instead, these data support the notion of a *true* perceptual effect, independent of semantic (labeling) processes. (Also see Campbell, Lewis, & Hunt, 1958, for a demonstration of contrast effects using an *absolute* and *extraexperimentally* anchored response language.)

These results have direct implications for communication. In many communication settings, the receiver's task is to infer the referent that the communicator had "in mind" when he constructed his message; thus, as in the method of single stimuli, the message (stimulus) must be placed into one of several categories (each category representing a different referent). If extreme contexts lead to perceptual contrast, by biasing the receiver's prior experience it should be possible to affect his choice from the set of potential referents, even if the available alternatives are reasonably familiar, unambiguous, and extraexperimentally anchored. In particular, a receiver who is provided with an *extreme* set of messages to be decoded (context) should, as a result, associate *neutral* messages with referents which *contrast* with this background of experience. In addition, if we assume that normally there is no "constant error" in communication, we are led to the further prediction that the contrast effects produced by extreme contexts should generally impair communicative accuracy (i.e., the receiver should be relatively unsuccessful in selecting the proper referent for a given message).

To test these hypotheses, subjects were presented with a series of 72 written messages, each describing an actor's portrayal of some emotional state. In one task, subjects were given the 11 photographs that had been used to elicit the various descriptions and were instructed to indicate the picture (referent) that was being described in each message. To assess the impact of contextual

factors upon the recipients' choices, some subjects were presented with descriptions of emotional states that were predominantly unpleasant, others were given mainly pleasant descriptions, and a third group responded to an unbiased mixture of both pleasant and unpleasant descriptions. In addition to these extreme (context) stimuli, each subject was also given 12 descriptions of pictures that were essentially neutral with respect to pleasantness-unpleasantness; these descriptions, which were presented intermittently, served as test stimuli. If biased experience produces perceptual contrast, then subjects who have mainly read descriptions of unpleasant emotions should interpret neutral descriptions as if their referents were relatively pleasant, as compared with subjects who have been primarily exposed to pleasant descriptions. Moreover, the accuracy of each subject's choices should be adversely affected when he is assigned to an extreme context condition.

METHOD

Messages

Thirty-one students from an introductory psychology course at the University of Michigan were recruited to participate in a study concerned with their "ability to communicate." The students were shown a series of 13 slides from the Frois-Wittmann series (2 were warm-ups) and were instructed to write their impressions of the "mood, feelings, or thoughts of the actor in each picture, or [to] give an indication of how you would feel if you were making that particular expression." The writers were further instructed to "do this in such a way as to convey enough information . . . so that someone else would know which picture you wrote about." The slides on which the descriptions were based were spaced at approximately equal intervals along the pleasant-unpleasant dimension; the following pictures were used: 4, 11, 22, 29, 36, 37, 47, 55, 66, 70, 72.

Three criteria were used in selecting messages from this overall pool:

1. The selected messages had to be readily classifiable with respect to pleasantness-unpleasantness.
2. Messages were excluded if they were too strongly focused on purely "physical" aspects of the stimuli (e.g., the fact that the actor's mouth was open), due to the unequivocal with which such messages might be linked to a given picture in the matching task.
3. Since the experiment required the use of 15 descriptions for each picture, an attempt was made

to select passages that were relatively diverse in content.

The following passages are typical of the descriptions that were used:

He is hilariously happy about a situation or joke or whatever. It is a laughing type of happiness, not just enjoyment or amusement. He could be taking a roller coaster and showing his happiness this way.

The actor is in a state of tension much like that experienced when watching a high diver posed to dive, but who has not leaped yet. Expectation also seems evident on his face.

This guy may be dying of thirst or else undergoing some minor pain. The expression is one of anguish but the suffering does not seem particularly intense.

Recipients

The recipients were paid college students, recruited at the University of Michigan and at Wayne State University. Subjects were tested in groups, ranging from 1 to 12; they were assigned to one of three distinct experimental tasks, henceforth identified as *match*, *rate*, or *parallel match*.

Matching task. This group of subjects ($N=72$) was instructed to match each passage with the picture that the writer was attempting to describe. That is, given 11 photographs randomly arranged on a sheet, they were to indicate the picture that had led to the writer's description.

Rating task. These subjects ($N=48$) were simply required to rate the degree of pleasantness-unpleasantness of each emotional state "as it would normally be experienced by the person being described." Ratings were made using an 11-point scale that ranged from *extremely unpleasant* (1) to *extremely pleasant* (11).

Parallel-Match task. In comparing the matching and rating tasks, it was clear that there were two important respects in which they differed: (a) The response alternatives (photographs) in the matching procedure seemed more explicit and thus less susceptible to redefinition than their counterparts in the rating task; (b) the matching instructions did not specify any particular dimension of the messages to which the subject should attend, while the rating instructions did. This suggested that if contrast effects were obtained in the rating procedure but failed to appear in the matching task, it might be difficult to interpret the results unequivocally. The parallel-match task ($N=48$), which may be viewed as a combination of the rating and matching procedures, was designed to guard against this possibility.

As in the rating task, subjects were instructed to gauge the pleasantness of the emotion underlying each of the descriptions. Then, rather than indicating their judgment in numerical form, from a series of 11 randomly arranged photographs subjects were to pick out the one picture that most nearly matched the description they had read, with respect to

pleasantness. Recipients were informed that, although none of the descriptions were based on the photographs which they were given, nevertheless each description had a matching picture in terms of the pleasantness of the underlying emotion; subjects were given the following Frois-Wittmann pictures from which to choose: 3, 5, 12, 16, 19, 28, 34, 48, 54, 57, 71.

Test Booklets

Within the rating and parallel-match tasks, two groups of subjects ($N=24$) were assigned to "extreme" context conditions; one group responded to a preponderance of unpleasant descriptions, and the other to mainly pleasant ones. These extreme descriptions were based on the 4 most pleasant and 4 least pleasant photographs in the set of 11. In the matching tasks there was, in addition, a "mixed" (or unbiased) context group, which was established by selecting context descriptions from both extremes of the pleasant-unpleasant continuum. Within each of the three tasks, subjects were randomly assigned to the various context groups; this was accomplished within each data-collection session by simply distributing a variety of test booklets representing the different context conditions.

In addition to the context items (there were 60 in each condition), all subjects responded to a common set of 12 test items; these were descriptions of emotional states that were essentially neutral with respect to pleasantness-unpleasantness. The neutral emotions occupied Positions 5, 6, and 7 when the 11 photographs of the total set were ranked on the basis of pleasantness-unpleasantness.

The descriptions were presented in mimeographed form with the test items interspersed among the context items. There were 12 pages in each booklet and six items per page; one description on each page was a test item (the test item always appeared as the third, fourth, fifth, or sixth item on the page), and the other five were context items.

All subjects were encouraged to respond to each description separately. They were warned that they might read many descriptions that seemingly match the same picture (or require the same rating, for subjects in the rate condition) or they might find that some pictures (ratings) might rarely, if ever, seem appropriate. In either of these cases, they were to be unconcerned and were simply to respond to each passage in its turn. These instructions were included in an effort to counteract the subject's possible expectation that the available response alternatives were to be used with roughly equal frequency.

To control for effects that might be attributable to differences between the various test and context items, a Latin square design was employed. In essence, within each of the main experimental conditions (e.g., rating task, pleasant context) there were 12 different arrangements of the experimental materials, so that each page of mimeographed descriptions (including both test and context items) appeared in each of the 12 possible ordinal positions. Two subjects responded to each of these 12 book-

lets. By combining the subgroups in each condition, it was thus possible to chart changes in response to a fixed pool of test items.

Dependent Variables

Scale value. The dependent variable for subjects in the rate condition was the numerical rating which they assigned to each test item. In the *match* and *parallel-match* conditions, the subject's choice of a given picture was converted into numerical form by assigning each picture a value between 1 and 11, based on its perceived position on the pleasant-unpleasant continuum, as derived from Schlosberg's (1952) norms. Schlosberg's data were based on a 9-point rating scale, which was converted to an 11-point scale by means of the formula: $Y = 1.25X - .25$, where Y represents the new (11-point) scale value, and X represents the values tabled in Schlosberg's norms.

Accuracy. For subjects assigned to the matching task, communicative accuracy was assessed by counting the frequency with which the subject selected the proper photograph (referent) in decoding the test items.

Experimental Design

Within each of the experimental tasks (*match*, *rate*, or *parallel match*), the experiment took the form of a mixed experimental design. There were two *between-subjects* variables associated with (a) context differences, and (b) subgroup differences based on the particular arrangement of pages in the experimental booklet, which was treated as a random effect. There was also one *within-subjects* variable, representing the successive test items. In evaluating the *scale value* data (see above), each of the experimental tasks was analyzed separately; to provide somewhat greater stability for the individual observations, the 12 test items were collapsed into 6 scores by summing responses to Items 1 and 2, 3 and 4, etc. A similar analysis was performed to evaluate the *accuracy* of the subject's performance in the matching task.

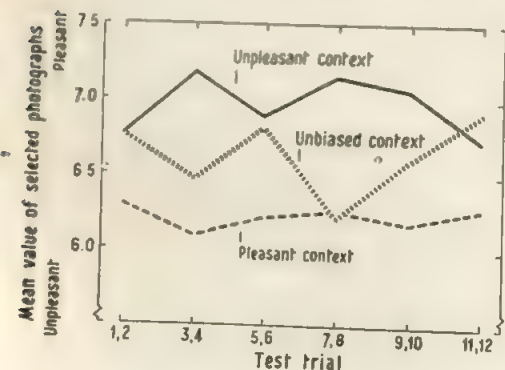


FIG. 2. The effects of extreme contexts on the subject's response in the matching task.

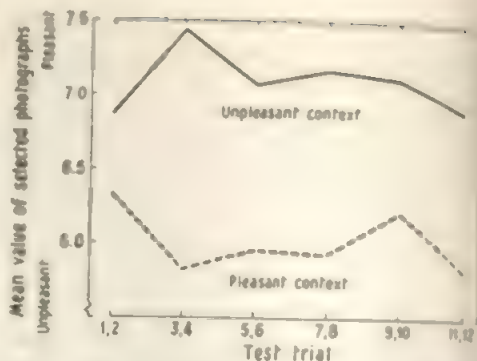


FIG. 3. The effects of extreme context on the subject's response in the parallel-match task.

RESULTS

Figures 2, 3, and 4 present the scale values obtained in the *match*, *parallel-match*, and *rating* tasks, respectively. Each figure shows the mean response elicited by pairs of successive test items; data are plotted separately for subjects assigned to different context conditions. The results show consistent evidence of a contrast effect. In all three tasks, subjects assigned to the pleasant context responded to the test items in less favorable terms than did those assigned to the unpleasant context. This trend appeared in every comparison of the pleasant and unpleasant context groups. Data from subjects assigned to an unbiased context within the matching task followed the expected pattern by generally falling between the two extreme groups.

To determine the reliability of these findings, separate analyses of variance were com-

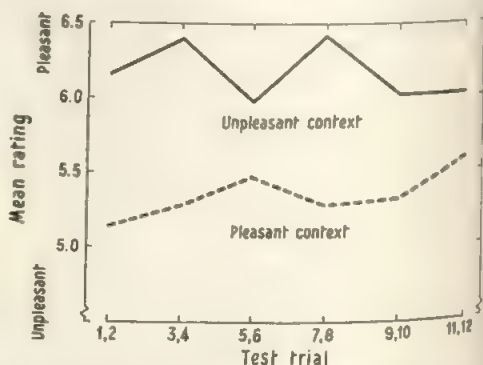


FIG. 4. The effects of extreme context on the subject's response in the rating task.

used for each task. In each case (see Table 1) the results attributable to context were significant when tested against a pooled error term based on the Context \times Booklet interaction, plus the variance due to subjects within groups.² Surprisingly, however, the obtained context effects appear to be reasonably stable throughout the experimental session, for the interactions between context and test trial were far from significant; thus, there is no evidence that context effects were built up gradually. Instead, it appears that in all three tasks the context which affects the subject's response to a given item is established within the few trials just preceding that item, for the effects of context are clearly apparent on the first two test descriptions when the subject has been exposed to 10 context items (or

fewer). Inspection of the data from Test Trial 1 *A* (Booklet) does not substantiate after this picture in the rating and parallel-match tasks, in the matching procedure. However, the difference between pleasant and unpleasant context groups is somewhat smaller on Trial 1 than on Trial 2.

The obtained interaction between test trials and booklets (see Table 1) is relatively devoid of systematic significance, despite its stability across tasks. This interaction simply represents the fact that the differences between the various subgroups of a given task-context combination (e.g. matching task, unpleasant context) were not constant from one test trial to the next; this is probably due to uncontrolled differences between the specific descriptions to which these subgroups were responding on any given test trial.

It is interesting to note that while the rating procedure resulted in significant contrast effects, the obtained *F* is far less striking than that obtained in the match and parallel-match tasks. Inspection of the summary tables indicates that this is mainly attributable to a large subjects within-groups error term, indeed, in the rating task, the error variance

TABLE 1
SUMMARY FOR ANALYSES OF VARIANCE FOR SCALE SCORES

Source	Task						
	Match*			Parallel match *		Rate	
	<i>df</i>	<i>MS</i>	<i>F</i>	<i>MS</i>	<i>F</i>	<i>MS</i>	<i>F</i>
Between Ss	47						
A (Context)	1	164.26	25.27****	338.44	49.41****	188.50	6.26**
B (Booklet)	11	9.19	1.29	7.14	1.20	7.06	
A \times B	11	5.10		8.78	1.47	22.69	
Ss within groups	24	7.14		5.97		53.53	
Pooled error	35	6.50		6.85		30.12	
Within Ss	240						
C (Trials)	5	1.06		2.16		1.48	
A \times C	5	3.10		5.89	1.36	4.53	1.13
B \times C	5	7.83	2.26*	10.23	2.43**	16.54	3.90***
A \times B \times C	55	3.74	1.08	4.58	1.09	3.55	
C \times Ss within groups	120	3.46		4.21		4.24	
Pooled error	175	3.54		4.33		4.02	

* To facilitate comparison with the parallel match and rating tasks, the analysis of the matching-task data is based only on data obtained in the two extreme contexts (pleasant and unpleasant). The results are unchanged when data from subjects in the neutral context are included.

** $p < .10$.

*** $p < .05$.

**** $p < .01$.

***** $p < .001$.

between subjects is greater by a factor of 4.7 than the *within-subjects* error term in the parallel form data and is 8.7 times greater than the equivalent term in the matching task (both $p < .01$). This inflated error term reflects the relative ambiguity of the response categories in the rating task when compared with the match and parallel match procedures. In essence, these results support the view that in a typical rating task the response alternatives are poorly defined, lending further credence to the contention that these alternatives may be particularly susceptible to redefinition when extreme contexts are introduced. Note however that the ambiguity here derives from *between-subjects* differences in the use of the rating categories. The fact that the *within-subjects* error term is roughly equal in the three experimental tasks suggests that the available response alternatives had comparable stability of meaning for the *individual subject* in all three cases.

Communicative accuracy. Table 2 presents the summary of an analysis performed on the accuracy data: the significant context effect is attributable to the fact that subjects assigned to the unbiased context were more successful in selecting the proper referent for the test items than were those in either the pleasant or unpleasant context groups. On the 12 neutral test descriptions, subjects in the unbiased condition showed a mean performance of 65% correct (7.7), in contrast to

those in the pleasant and unpleasant context groups who averaged 51% (6.1) and 68% (6.8), respectively. Comparison of the means by the method of orthogonal contrasts (Hays, 1963, p. 466) shows a significant gap between the unbiased group mean and the average obtained from a pooling of the two extreme context groups; the unbiased group is also significantly more accurate than each of the biased groups considered singly.

As in Table 1, the interaction ($B \times C$) between test trials and the experimental groups which were established as part of a Latin square is of minor significance; it reflects the fact that the differences between subgroups varied from one trial to the next, probably because the test items varied in difficulty. On a given trial, one subgroup might be responding to a difficult item, and they might perform poorly relative to other subgroups faced with simpler tasks; on another trial, however, the first group might show excellent performance, if its test item was an easy one.

DISCUSSION

The marked contrast effects obtained in the matching task suggest that in everyday communication the recipient's understanding of a message, that is, the referent which he assumes to have guided the communication output, may be significantly altered by prior exposure to an extreme sample of messages from the same domain. Note, however, that despite the present efforts to select response alternatives in the matching task that would be unequivocal in meaning and not susceptible to redefinition we have no *direct* evidence that this attempt was successful; hence it is conceivable that the obtained contrast effects result from a displacement of the subjective meanings associated with the response alternatives (photographs), rather than from a

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE FOR
COMMUNICATIVE ACCURACY

Source	df	MS	F
Between Ss	71		
A (Context)	2	2.39	4.35*
B (Booklet)	11	.59	1.16
A \times B	22	.61	1.20
Ss within groups	36	.51	
Pooled error	58	.55	
Within Ss	360		
C (Trials)	5	.59	1.38
A \times C	10	.43	1.00
B \times C	55	.84	1.91***
A \times B \times C	110	.42	
C \times Ss within groups	180	.44	
Pooled error	290	.43	

* $p < .025$.

*** $p < .01$.

* The superior performance exhibited by unbiased subjects in response to the test items also appears in the context items, although the group differences here are somewhat diminished. Thus, in response to descriptions of pleasant emotions, the neutral group is correct 60% of the time, while subjects in the pleasant context show a 57% hit rate; similarly, the neutral group succeeds in picking the correct referent for 64% of the unpleasant items, while subjects in the unpleasant group are correct on 61%.

accuracy of the responses elicited by the context. In other cases, however, it appears that the introduction of a biased context has a significant impact upon the percentage of correct responses associated with a verbal statement, and may, moreover, adversely affect communicative accuracy (i.e., the receiver may be relatively unsuccessful in recovering the speaker's intended meaning).

The matching task data are important in demonstrating that contrast effects may be obtained even though the experimenter does not make out a particular dimension for the subject to focus on. Apparently these effects may result without any special "tuning," although it seems likely that the relative salience of the pleasantness dimension in descriptions of emotion may have been important here. Further research will be required to determine whether these effects can be obtained in a "free" setting when the contextual stimuli are biased with respect to a less salient dimension (e.g., activity-passivity).

Communicative accuracy. The relatively poor accuracy shown by subjects assigned to the extreme contexts seems closely related to the displacement effects noted earlier. In assessing referents for the test messages, these subjects were less likely to select photographs from the midregion of the pleasant-unpleasant continuum (photographs in Rank Positions 1-7) than were those in the unbiased condition. This reduction in midregion choices, which were the true referents for the test messages, resulted from an increased frequency of "contrast" responses, with virtually no compensating shift from the "context side" of the continuum into the midscale region. For example, subjects assigned to the pleasant context selected unpleasant photographs more frequently than did those in the unbiased condition; these two groups were similar, however, in their choice of pleasant photographs.

In an effort to document further the impact of these context-induced displacements on communicative accuracy, Table 3 presents an analysis which separates incorrect responses into two types: those in which the selected referent was less pleasant than the true referent, and those in which it was more pleasant. As shown previously, subjects in the unbiased

TABLE 3

Percentages of Errors in a Matching Task

Context	More Pleasant	Less Pleasant	Total
Pleasant	24.7%	11.3%	36%
Unbiased	24.1%	10.9%	35%
Unpleasant	22.7%	14.5%	37.2%

Percentages of errors in a matching task, by context.

context exhibited fewer errors than those in the biased conditions. The inference that the relevant group of test messages of a response here is that pleasant errors are more than twice as common as "unpleasant" errors (14% versus 11%). This effect seems worthy of further study since it is a systematic error that may be present in many communication settings.

Using the error distribution obtained in the unbiased group as a base line, it is clear that the reduced accuracy of the biased subjects is due to increase in contrast errors. That is, subjects in the pleasant context show an increased tendency to choose photographs less pleasant than the true referents, and there is an analogous increase of pleasant errors among subjects assigned to the unpleasant context. Note, however, the relatively constant percentage of errors resulting from choices on the context side. Thus, among subjects assigned to the pleasant context, the percentage of error responses more pleasant than the true referent is nearly equal to that obtained in the unbiased group. Similarly, subjects in the unpleasant context show almost the same frequency of unpleasant errors as was observed in the baseline condition. These results follow quite directly from the displacement effects discussed above.

REFERENCES

- CAMPBELL, D. T., HUNT, W. A., & LEWIS, N. A. The effect of assimilation and contrast in judgment of clinical materials. *American Journal of Psychology*, 1957, 70, 197-212.
- CAMPBELL, D. T., HUNT, W. A., & LEWIS, N. A. The relative susceptibility of two rating scales to disturbances resulting from shifts in stimulus context. *Journal of Applied Psychology*, 1958, 42, 213-217.

- CAMPBELL, D. T., LEWIS, N. A., & HUNT, W. A. Context effects with judgmental language that is absolute, extensive, and extra-experimentally anchored. *Journal of Experimental Psychology*, 1958, **55**, 220-228.
- HAYS, W. L. *Statistics for psychologists*. New York: Holt, Rinehart & Winston, 1963.
- HINTZ, R. K. The effect of remote anchoring points upon the judgments of lifted weights. *Journal of Experimental Psychology*, 1950, **40**, 584-591.
- HILSON, H. *Adaptation-level theory: An experimental and systematic approach to behavior*. New York: Harper & Row, 1964.
- KRANTZ, D. L., & CAMPBELL, D. T. Separating perceptual and linguistic effects of context shifts upon absolute judgments. *Journal of Experimental Psychology*, 1961, **62**, 35-42.
- SCHLOSSBERG, H. Ratings on pleasantness, unpleasantness and rejection-attention on a scale of 1 to 10. *Journal of Experimental Psychology*, 1958, **55**, 229-237.
- STEVENS, S. S. Adaptation level vs. the reference point in judgment. *American Journal of Psychology*, 1962, **71**, 663-647.
- VOLKMANN, J. Scales of judgment and their applications for social psychology. In J. H. Keesen & M. Sherif (Eds.), *Social psychology at the crossroads*. New York: Harper, 1951. Pp. 27-44.
- UPSHAW, H. S. Own attitude as an anchor in appearing intervals. *Journal of Abnormal and Social Psychology*, 1962, **64**, 85-96.

(Received May 26, 1966)

ACQUISITION OF A HOSTILE ATTITUDE AND ITS RELATIONSHIP TO AGGRESSIVE BEHAVIOR¹

CLEMENS A. LOEW²

University of Iowa

During the 1st phase of an experiment, $\frac{1}{2}$ of the Ss were trained to deliver aggressive words (hostile group) and the other $\frac{1}{2}$ were trained to deliver non-aggressive words (nonhostile group) from sets of 4 different words which they read aloud to S. In the 2nd phase, all Ss externally delivered electric shock to another "S" while carrying out a training procedure. In each of the above groups delivered shock under an interdependent condition designed to create frustration, and the other $\frac{1}{2}$ under an independent condition. The results indicate that differential training of aggressive verbalization was effective in producing differential physical aggression intensity (but not duration). Support was not found for the differential effect of the frustrating conditions. The results offer support for the view that hostile responses mediate overt aggressive behavior.

In the study of aggressive behavior, some attention has been given to the variables affecting the acquisition of aggressive verbal responses and their effect on other aggressive behavior. Two trends are of interest in the present context. First, verbal reinforcement techniques have been found to produce increments in the occurrence of hostile words (Williams, 1964). Second, a recent development in the study of aggressive behavior in experimental laboratory settings is the utilization of the "aggression machine" (Buss, 1961). This technique provides subjects with the opportunity to administer electric shocks of varying intensity to another person and permits study of the variables that affect this overt aggressive activity. The purpose of the present study is to examine the effect of training of verbal aggressive responses on subsequent exposure to the aggression machine in certain relevant social contexts.

¹This study is based upon a dissertation presented to the Department of Psychology of the University of Iowa in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The writer wishes to express his appreciation to Milton E. Rosenbaum for his guidance and personal encouragement throughout the course of this investigation. The assistance of the author's doctoral committee members, Sidney J. Arenson, Donald J. Kiesler, Leonard D. Eron, and Jan Loney, is also gratefully acknowledged.

²Now at the Postgraduate Center for Mental Health, New York City.

Although the learning, extinction, and generalization of verbal behaviors have received attention for several decades (Williams, 1964), Buss and Durkee (1958) were probably the first to demonstrate the effectiveness of the experimenter's verbal reinforcement in increasing the frequency of usage of hostile words in sentence construction. Zedek (1959), employing a word-association test, also found support for the learning of hostile responses as a function of verbal reinforcement. Binder (1957) and Ferguson and Buss (1960) have shown that effectiveness of reinforcement of hostile responses varies with the experimenter's role characteristics.

In research pertaining to the study of physical aggression employing shock delivery as the mode of aggression, a variety of variables have been examined. For example, Buss (1961) investigated the effect of sex differences on the intensity of shock administration to an accomplice, Walters and Thomas (1963) examined the effects of viewing aggressive films on subsequent selection of shock levels, and Staples and Walters (1964) found that the effectiveness of reinforcement on shock delivery is a function of initial levels of inhibition of aggressive behavior. These studies provide support for the usefulness of this technique as an objective and direct study of some of the determinants of physical aggression.

The purpose of the present study is to examine the relationship between the learning of aggressive verbal responses and the manifestation of subsequent related responses, particularly those of physical aggression. Lovaas (1961) has been unique in examining the effects of strengthening verbal responses on nonverbal aggressive responses. He showed that children aged 3-5 and 4-7 who were reinforced for aggressive verbal responses (e.g., "bad doll") exhibited a higher proportion of subsequent doll striking than those reinforced for nonaggressive words (e.g., "good doll").

Verbal responses that are not delivered as noxious stimuli may be classed as expressions of hostility, and hostility may be viewed as an attitude. Rhine (1958) defines attitude as a concept or an implicit mediating response with an evaluative dimension. According to this view, an attitude is an internal event which serves as a mediator in the sense that once such an implicit cue-producing response has been established, other responses, internal or overt, can be conditioned to it.

Research dealing with attitudes as mediators of evaluative behavior has not been extensive, but studies such as those by Eisman (1955), Staats and Staats (1958), Rhine and Silun (1958), and Radtke (1963) do provide sufficient evidence for the conditionability and learning of attitudes and for their mediational qualities.

Compatible with the present approach to attitude is Buss' (1961) definition of hostility which he views as an attitudinal event. Hostility is an "implicit verbal response involving negative feelings (ill-will) and negative evaluations of people and events [p. 12]." When verbalized, hostile responses take the form of negative labels of the kind used in derogatory statements such as "I hate him" or "I despise him." Aggression, on the other hand, is considered an overt response—"a response that delivers a noxious stimulus to another organism [Buss, 1961, p. 1]."³ These definitions differ somewhat from others, such as those given by Dollard, Doob, Miller, Mowrer, and Sears (1939) and

Berkowitz (1958, 1962) who use hostility and aggression interchangeably.

According to the approach taken in this study, the training of aggressive verbalizations, which are assumed to be one form of hostile attitude, strengthens the hostile mediational responses whose cues, together with their evaluative elements, become a discriminative stimulus for aggression to occur. In this sense the verbal conditioning studies employing hostile verbal materials reported above have demonstrated the development of hostile attitude.

It is likely that attitudinal mediation of aggressive behavior occurs only under certain conditions of instigation to aggression. Among instigators capable of evoking a hostile attitude are frustrating conditions produced by another person. One specific frustration condition that may elicit aggression is one in which two individuals, A and B, work together interdependently on a cooperative task, and B's failure prevents A from successful completion of the task. Since B's errors block rewards for A, B may be considered a frustrator and elicit hostility (Jones & deCharms, 1957). This may be contrasted to a comparable situation in which A and B also work together, but where B's failure does not prevent rewards for A. Since it is possible for A to obtain reward independently, then B's errors are less likely to produce frustration. Thus, interdependence as opposed to independence should create more frustration when another fails. As a consequence interdependence should also lead to greater likelihood of aggressiveness.

In sum, it is the strength and dominance of a hostile attitude in conjunction with the presence of an instigation that determines the strength of overt aggression. In the present study the hypothesis to be examined is that the acquisition of aggressive verbal responses, as opposed to nonaggressive verbal responses, leads to differential physical aggressive behavior depending on the instigation characteristics of the social conditions. Specifically, the primary interest was in studying the effect of training of aggressive and nonaggressive verbal responses on subsequent intensity and duration of shock delivery

³This may include motor or vocal behavior. Hitting someone or saying "I hate you" are both aggressive responses,

under interdependent and independent social conditions.

METHOD

Subjects

The subjects were 80 college males enrolled in a course in introductory psychology. There were two levels of the attitude-training variable and two levels of the cooperation variable. The subjects were randomly assigned to one of four groups with 20 subjects in each: hostile interdependent, nonhostile interdependent, hostile independent, and nonhostile independent.

Apparatus and Materials

Phase I: Attitude acquisition. Materials for this part involved four different classes of words: 60 aggressive verbs, 60 neutral verbs, 60 nouns pertaining to household items, and 60 nouns pertaining to nature. From these, 60 sets of four different words were randomly compiled. Each set contained one word from each of the four classes, and the order of appearance of the classes in each set was randomly assigned. These sets were typed on cards and were designed for the experimenter to read aloud to the subjects.

Phase II: Expression of aggression. In the same experimental room in which Phase I was conducted a modified Buss (1961) type aggression machine was placed on a table in front of the subject. The subject's panel contained a row of 10 buttons, and above each button was a light which flashed for the duration of the button press. To the left of these buttons was a red light that signaled a team or other "subject" error. About 6 inches directly above that light was a green light that signaled subjects or team accuracy. In the center of the panel and about 1 foot above the row of buttons was a white light which when turned on signaled the onset and duration of each trial. A wristband containing two electrodes was connected to the first two buttons. Two intensities of electric shock, 24 volts and 30 volts, could be administered by pressing the first and second button, respectively. These buttons were used for demonstration purposes only. The other eight buttons were not wired for shock.

In front and to the right of the panel was a microphone which led to a speaker placed in the adjacent room (experimenter's control room). Also available to the subject were 14 5 × 7 cards. Each card had four different words printed on it in capital letters. The words were numbered 1, 2, 3, and 4. On each card one of the four words pertained to some sport, and it was printed in different positions on each of the 14 cards. The other words on the cards were randomly selected and did not pertain to sports or any other specific class. In addition, the subject had a list of 14 different orders of presentations, one order for each set of four words on a card. This material was used by the subject in Phase II.

The recording apparatus, located in the adjacent

control room, consisted of a panel which depicted a row of 10 lights which corresponded to the 10 shock buttons of the aggression machine. When a certain shock button was pressed by the subject, its corresponding light was turned on, thus permitting recording of the level (intensity) of shock. In the center below these lights was a "start" switch which turned on the white light on the subject's panel and signaled the start of a trial. The panel also contained one switch for signaling a correct trial to the subject and one for signaling an error. There was a Hunter decade clock counter hooked up via relay to the shock buttons. Depression of any shock button by the subject activated the timer and release of the button stopped it. The experimenter reset the clock after every trial. This provided a time measure of the duration of each button press. In addition, headphones were available to allow the experimenter to hear the subject perform his task.

Design and Procedure

There were two categories on each of the two independent variables: type of attitude, hostile or nonhostile; and type of cooperative condition, interdependent or independent.

Phase I. The experimenter read a set of four words at a time, and the subject was asked to choose one of them and repeat it aloud. The subject was informed that the experimenter would tell him when a correct word was chosen. One group (hostile) was reinforced for choosing aggressive words, and one group (nonhostile) for choosing nonaggressive words pertaining to nature. Reinforcement consisted of the experimenter saying "correct" whenever the subject chose an appropriate word. Training continued until the subject reached the performance criterion of 20 successive correct responses. In the typical verbal conditioning studies involving sentence construction, failure to achieve criterion occurs frequently. The complexity of this verbal training task was drastically reduced to enable subjects to learn to a specified performance criterion.

Phase II. The experimenter went into the control room, and the subject was left alone in the experimental room. One-half of each of the above two groups ostensibly delivered "shock" to another "subject" (actually nonexistent) under the interdependent condition and the other half under the independent condition. The following information was available to each subject: The experimenter was interested in the proficiency of college students as experimenters, and the subject was to train another "subject" in the next room to select words pertaining to sports. He was to read into the microphone on each trial a set of four words presented on a card. The other "subject" was to choose a word from the four that the subject presented by flipping one of the four switches on his panel. If a correct response occurred, the green light was illuminated and the subject was to do nothing. If an error occurred, the red light went on, and in this case the subject was to deliver shock to the other "subject" by pushing 1 of the 10 buttons. There were 14 trials for all

subjects, 10 of which involved errors by the "subject" and hence the administration of shock. On Trials 4, 7, 11, and 14 correct words were chosen, and the subject was not required to push any buttons.

Instructions for the interdependent condition indicated that the evaluation of the subject's performance as an "experimenter" was based on the combined performance of the subject and his "subject" as a team. The subject was also informed that the recording apparatus was located in a third room and that it was connected in a manner which only allowed recording of the final group product. It was indicated that since the experimenter was unable to record individual performance, the subject's efficiency as an "experimenter" depended on the success of his "subject." Instructions for the independent condition informed the subject that the evaluation of his performance as an "experimenter" was based on his own individual performance. It was indicated that since the experimenter was able to record both individual performances, the subject's efficiency as an "experimenter" did not depend on the other "subject's" performance. ϵ

In order to achieve maximum contiguity between verbal training and motor expression, the chronological sequence of procedures for all subjects was as follows: First, the subject was instructed about his task in Phase II. The subject also received a demonstration of mild shock from Buttons 1 and 2 and was informed that the other "subject" did not have the power to retaliate. Second, the experimenter trained the subject to use aggressive (or nonaggressive) words (Phase I). Third, the subject was refreshed about the procedure he was to follow in Phase II. Fourth, the subject was informed about the cooperative conditions. Fifth, according to the instructions given to the subject earlier, he performed his task as an "experimenter" (Phase II).

The following deceptions were employed by the experimenter to facilitate the belief that another "subject" was present in the adjacent room. The sign-up sheet had two spaces for names for each of the times for which the experimenter was scheduled. One of these spaces was filled with a fictitious name written by the experimenter, and the other remained open for a subject to sign. Signs with the same experiment number were placed on the doors of the two adjacent rooms. Visible wires were connected between the two adjacent rooms and a third room. In addition, when the subject entered the experimental room, he was told that the experimenter was going to bring the other "subject" to the next room and give him his instructions first. The experimenter then left the room and returned after several minutes.

RESULTS

Phase I

To determine if differences between the speed of learning of aggressive and nonaggressive

words are present, a two-tailed t test was performed on the number of total trials required to learn these two kinds of words to the performance criterion. The results indicated that the hostile groups learned the aggressive words ($M = 24.35$, $SD = 5.67$) in significantly fewer trials than the nonhostile groups ($M = 30.27$, $SD = 10.93$) learned the nonaggressive words ($t = 3.02$, $df = 78$, $p < .01$). Thus, the choice of aggressive words was easier to learn than the choice of nonaggressive words.

Phase II

To evaluate the effects of verbal training (Phase I) upon subsequent motor responses, two response measures were used: shock level (intensity) employed by the subject and duration (in seconds) of the button press. The mean shock levels for the four groups for two blocks of five trials and for the total 10 trials are presented in Table 1. The results of a Type III (Lindquist, 1953) analysis of variance of these data showed that the hostile groups delivered shocks of significantly greater intensity than the nonhostile groups ($F = 19.13$, $df = 1, 76$, $p < .001$).

A significant Trial \times Training interaction was obtained ($F = 5.83$, $df = 1, 76$, $p < .025$) indicating that the hostile and nonhostile groups differed in shock increments from the first to the second block of trials. A further analysis of the simple effects using

TABLE 1
MEAN SHOCK LEVELS IN TWO SUCCESSIVE BLOCKS OF
FIVE TRIALS AND FOR THE TOTAL
TEN TRIALS*

Group	Trials					
	1-5		6-10		1-10	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Hostile inter-dependent	4.07	1.54	5.77	1.85	4.92	1.58
Nonhostile inter-dependent	3.03	1.72	3.98	1.90	3.51	1.76
Hostile independent	4.31	1.34	5.84	2.01	5.05	1.61
Nonhostile independent	2.73	1.24	3.78	1.87	3.23	1.47

* The means are based on the means of individual scores over trials.

the within-subject error term showed that over trials there was a significant increase in shock levels for both the hostile groups ($F = 80.25$, $df = 1, 76$, $p < .001$) and non-hostile groups ($F = 33.33$, $df = 1, 76$, $p < .001$). Thus since there was a significant interaction and since from inspection of Figure 1 it is clear that both groups increase shock intensity, it is concluded that the hostile group increased more than the non-hostile group. None of the tests involving the cooperation conditions were significant.

The mean durations of shock are presented in Table 2. For the four groups the trend for duration is similar to that of intensity, but a Type III analysis of variance yielded no significant results.

Because of the possibility of a relationship between trials to acquisition (Phase I) and subsequent shock level employed by the subject (Phase II), an overall Pearson product-moment correlation was obtained ($r = -.185$, $p > .05$). The results indicate that there is no significant relationship between the number of trials to learn the aggressive or

TABLE 2

MEAN DURATION (IN SECONDS) OF SHOCK IN TWO SUCCESSIVE BLOCKS OF FIVE TRIALS AND FOR THE TOTAL TEN TRIALS*

Group	Trials					
	1-5		6-10		1-10	
	M	SD	M	SD	M	SD
Hostile inter-dependent	.74	.52	.88	.63	.81	.57
Nonhostile inter-dependent	.66	.42	.79	.51	.73	.45
Hostile independent	.74	.39	.84	.52	.79	.43
Nonhostile independent	.56	.33	.59	.32	.58	.32

* The means are based on the means of individual scores over trials.

nonaggressive words and the level of shock employed.

DISCUSSION

The following is a summary of procedures employed in the experiment. First, subjects were trained to choose and speak aggressive or nature words to the performance criterion of 20 successive correct trials (Phase I). Then subjects were exposed to an aggression machine which provided an opportunity for the subject to deliver shock to another "subject" in the context of a learning experiment (Phase II). The present study sought to examine the effect of training of aggressive or nonaggressive verbal responses on subsequent intensity and duration of shock delivery under interdependent and independent social conditions.

Several aspects of the results require discussion. First, the results showed that the choice of aggressive words was easier to learn than the choice of nature words. This difference may be attributed to an apparently more homogeneous sample of aggressive as compared to the sample of nature words, which include words pertaining to geography, botany, and meteorology. Consequently, it appears that the aggressive words were more discriminable and thus easier to learn.

Second, the major and most significant finding was that an increase in subsequent physical aggression (intensity but not dura-

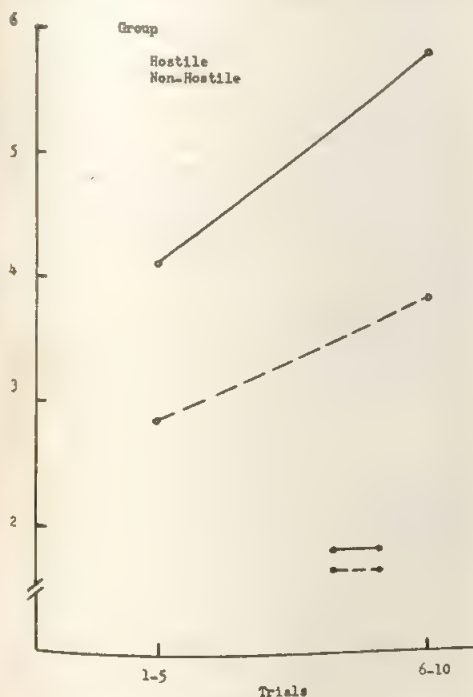


FIG. 1. Mean shock levels in successive blocks of five trials.

tion) was brought about by the training of aggressive verbalizations. The possibility that there was a relationship between ease of learning words (Phase I) and subsequent aggression (Phase II) was rejected on the basis of a small and nonsignificant correlation ($r = -.185$) between trials to acquisition and intensity of aggression. Thus, given the opportunity to aggress in a situation where retaliation is unlikely and where aggression is sanctioned, persons trained to be hostile will manifest greater aggression than those not trained to adopt this attitude. Moreover, it was found that hostile subjects increase their aggression over trials.

One distinctive feature of these findings is the occurrence of a transition from internal, mediational events to motor activity as a function of the reinforcement of selection and repetition of a class of words. It would appear that differential training of verbal responses produced an ability to discriminate among the quantitative designations associated with a row of 10 buttons. Not only did subjects trained for aggressive verbalizations choose higher shock levels, but they also shifted to a higher range as trials progressed (see Table 1).

These findings point to another important quality. If it can be assumed that subjects have inhibitions against aggressing, particularly delivering electric shock to another person, then the training of aggressive verbalizations must have been particularly effective in bringing about physical aggression. For all subjects, the use of shock was phrased as an instrumental response that was designed to improve the performance of the other "subject." These instructions could then be expected to reduce inhibitions for all subjects and not only the subjects trained to choose aggressive words. Furthermore, it might have been expected that since subjects vary in initial inhibition, verbal training should not have a uniform effect on subsequent aggression. That is, large variability in physical aggression should also be present. Staples and Walters (1964) found that actual reinforcement of shock was effective in producing pretest to posttest increase only in subjects with low inhibition. Inspection of Table 1 of the present findings, however, shows that

standard deviations are small and uniform for the aggression measure.

According to the present approach, it is assumed that physical aggression occurred because the mediational, fractional component of the overt response was previously strengthened through the reinforcement of verbal aggressive response. The finding that previous reinforcement history, and not frustration, was sufficient in producing aggression is contrary to the frustration-aggression hypothesis (Dollard et al., 1939). That hypothesis stated that aggression is primarily if not always caused by frustration.

Inspection of Table 2 indicates that the trend of the means for the duration of aggression for the aggressive word ($M = .80$ second) and nature word group ($M = .65$ second) is consistent with the intensity data. Failure to find differences in duration may be explained by the variability that was obtained for this measure (see Table 2). Although uniform instructions and demonstrations involving the utilization of shock emphasized the intensity of shock employed, no reference was made to the duration of shock delivery. Consequently, subjects were free to vary in the degree to which length of shock delivery was utilized for expressing aggression. This suggests that in conjunction with the training of aggressive verbalizations some direction or guidance regarding the mode of aggression is necessary to bring about increments in specific aspects of physical aggression. Lack of such guidance may produce variability in the selection of the modes of aggression.

Third, no support was found for the differential effect of the interdependent-independent social conditions on physical aggression in either mode of aggression. This suggests that regardless of social frustration conditions, hostile subjects aggress more than nonhostile subjects. It is possible, however, that the present interdependent condition was relatively ineffective in creating an instigation. It is possible that the effects of information obtained from the uniform instructions given to all subjects offset the effects of the specific instructions designed to produce frustration in the interdependent group. It will be recalled that all subjects were informed

that the purpose of the experiment was to study the proficiency of college students as experimenters, to study the effects of shock on learning, and that the subject's task was to train another "subject." Differential instructions for the interdependent and independent group referred to the experimenter's ability to evaluate the subject on a group or individual basis. It may be that under both social conditions subjects attended more to their task, that is, to train another "subject" by the use of shock, than to the experimenter's evaluation of their performance.

It is probable that the mere assessment of their proficiency by the experimenter was not a sufficiently desirable reward. Further investigation should consider a more definite reward, such as monetary reward as used by Jones and deCharms (1957), with the possibility that interference with reward will have stronger frustrating consequences.

Thus, while the present findings indicate that verbal reinforcement history affects subsequent physical aggression, the effects of the frustrating social conditions employed in this experiment are not clear.

REFERENCES

- BERKOWITZ, L. The expression and reduction of hostility. *Psychological Bulletin*, 1958, 55, 257-283.
- BERKOWITZ, L. *Aggression: A social psychological analysis*. New York: McGraw-Hill, 1962.
- BINDER, A., McCONNELL, D., & SJOHOLN, N. A. Verbal conditioning as a function of experimenter characteristics. *Journal of Abnormal and Social Psychology*, 1957, 55, 309-314.
- BUSS, A. H. *The psychology of aggression*. New York: Wiley, 1961.
- BUSS, A. H., & DURKEE, A. Conditioning of hostile verbalizations in a situation resembling a clinical interview. *Journal of Consulting Psychology*, 1958, 22, 415-418.
- DOLLARD, J., DOOS, L. W., MILLER, N. E., MOWLER, O. H., & SEARS, R. R. *Frustration and aggression*. New Haven: Yale University Press, 1939.
- EISMAN, B. S. Attitude formation: The development of a color preference response through mediated generalization. *Journal of Abnormal and Social Psychology*, 1955, 50, 321-326.
- FERGUSON, D. C., & BURN, A. H. Operant conditioning of hostile verbs in relation to experimenter and subject characteristics. *Journal of Consulting Psychology*, 1960, 24, 324-327.
- JONES, E. E., & DECHARMS, R. Changes in social perception as a function of the personal relevance of behavior. *Sociometry*, 1957, 20, 75-85.
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
- LOVAAS, I. O. Interaction between verbal and non-verbal behavior. *Child Development*, 1961, 32, 329-336.
- RADTKE, R. C. An S-R approach to attitude intensity. Unpublished doctoral dissertation, State University of Iowa, 1963.
- RHINE, R. J. A concept-formation approach to attitude acquisition. *Psychological Review*, 1958, 65, 362-370.
- RHINE, R. J., & SILUN, B. A. Acquisition and change of a concept attitude as a function of consistency of reinforcement. *Journal of Experimental Psychology*, 1958, 55, 524-529.
- STAATS, A. W., & STAATS, C. K. Attitudes established by classical conditioning. *Journal of Abnormal and Social Psychology*, 1958, 57, 37-40.
- STAPLES, F. R., & WALTERS, R. H. Influence of positive reinforcement of aggression on subjects differing in initial aggression level. *Journal of Consulting Psychology*, 1964, 28, 547-552.
- WALTERS, R. H., & THOMAS, E. L. Enhancement of punitiveness by visual and audiovisual displays. *Canadian Journal of Psychology*, 1963, 17, 244-255.
- WILLIAMS, J. H. Conditioning of verbalization: A review. *Psychological Bulletin*, 1964, 62, 383-394.
- ZEDEK, M. E. The conditioning of verbal behavior with negative cultural connotations. *Journal of Personality*, 1959, 27, 477-486.

(Received June 2, 1966)

BRIEF ARTICLES

A STUDY OF THE DROPOUT RATES IN LONGITUDINAL RESEARCH ON AGING AND THE PREDICTION OF DEATH¹

KLAUS F. RIEGEL, RUTH M. RIEGEL, AND GÜNTHER MEYER

University of Michigan

Intelligence, verbal abilities, attitudes, interests, and social conditions of 380 Ss above 55 yr. of age were measured. Five yr. later Ss were retested. Some refused to cooperate again, and others had died or become ill. Retested Ss differed significantly from the total group, but in particular from the other subgroups. The prediction of death on the basis of sociopsychological variables was more successful for Ss below than above 65 yr. of age. It was concluded that developmental trends are based on increasingly biased samples, that previous studies have underestimated the amount of attrition, and that nonsurvivors under 65 yr. form a sociopsychological subgroup of different characteristics than survivors.

During recent years, an increasing number of longitudinal studies of adult and aged subjects have been reported in the psychological literature. Most of these studies are based on special subsamples of the population and have applied restricted sets of psychological measures only. Owens (1953) retested the intelligence of a group of superior adults, and Bayley and Oden (1955) reported on the longitudinal analysis of Terman's subjects. At the other end of the continuum, Kaplan (1943, 1956) and Bell and Zubek (1960) retested mentally inferior persons. In both cases it would be inappropriate to generalize the reported findings since there is, most likely, an interaction between the rate of change and the original level of functioning. This has been emphasized by Foster and Taylor (1920), Jones and Conrad (1933), and Miles (1933).

More recently, Jarvik, Kallman, and Falek (1962), Jarvik, Kallman, Falek, and Klaber (1957), Jarvik and Falek (1963), and Falek, Kallman, Lorge, and Jarvik (1960), in a number of joint publications, reported on a longitudinal investigation of intellectual functioning and longevity of senescent twins. Kleemeier (1962) and Kaplan, Rumbaugh, Mitchell, and Thomas (1963) retested the intelligence of residents in a home for the aged after various time intervals, and observed sudden and marked performance decrements preceding the death of the subjects. Among the best-matched samples of the aging popula-

tion were those by Berkowitz and Green (1963), Eisdorfer (1963), and by Schaie and Strother (1964). The former two studies were restricted to the measurement of intelligence, while the latter also included some attitudinal scales applied to stratified samples ranging from 20 to 70 years of age.

In most of these studies, developmental psychologists have never questioned the superiority of the longitudinal over the cross-sectional design. Only very recently has the complementary nature of both strategies been recognized. In particular, Schaie (1959, 1965) and Schaie and Strother (1964) have provided a thorough discussion of the experimental strategies in gerontological research, in emphasizing more general designs in comparison with which the two traditional approaches are merely specialized and incomplete cases.

In a more concrete sense, the limitations of the traditional research strategies have long been recognized. In their cross-sectional study, Jones, Conrad, and Horn (1928) analyzed the performance of subjects who had originally refused to participate in their study, thus biasing the sample. More recently, Sussman (1964), Damon (1965), and Rose (1965) discussed the representativeness of samples in longitudinal research primarily concerned with the health of the subjects.

Because of systematic factors, of which sickness and death are the most obvious, the problem of sample bias is equally important for longitudinal and cross-sectional studies in psychological gerontology. Indeed, if systematic selec-

¹ This study has been aided by the Foundations' Fund for Research in Psychiatry, New Haven, Connecticut. The analysis has been completed at the Computing Center, the University of Michigan.

by dropout factors (such as selective death rates) can be detected, the concept of psychological development based on observed trends would itself be seriously challenged, because then cross-sectional or longitudinal research would represent averages for sets of systematically biased age samples only; therefore any inferences about general developmental trends would be questionable.

The present analysis has been undertaken primarily to determine the psychological characteristics of subjects who either did not survive the time period between the two testings, were too ill, or refused to be retested. Since significant differences between the subgroups were detected, attempts have also been made to predict the ensuing death of the subjects on the basis of sociopsychological factors. The present analysis is restricted to major comparisons between the four subgroups mentioned. (In a supplementary technical report—Riegel, Riegel, & Meyer, 1967—statistical details on the 43 variables and five age groups at both testings have been provided, and differences in the distributions of test scores and the effect of multiple testing have been discussed.)

METHOD AND PROCEDURE

Subjects

The present analysis is based on the results of the first testing of a study on sociopsychological factors of aging, conducted in Germany in 1956-57 and a retest study in 1961-62.

The original sample consisted of 190 females and 190 males. These cases were drawn from a group of about 500 subjects, and were subdivided into five age levels of 38 females and 38 males each. The five age levels were 55-59, 60-64, 65-69, 70-74, and over 75 years (average age = 79.0 years). Aside from controlling for age and sex, each age level was matched against census statistics on the following criteria: occupation, source of income, marital status, refugees versus nonrefugees, and religious affiliation. The samples can be regarded as representative for the population of northern Germany. Fuller descriptions of the samples and the procedures are given elsewhere (Riegel & Riegel, 1959; Riegel, Riegel, & Skiba, 1962).

At the time of the second testing all subjects had moved into the next higher age levels. Of the 380 persons originally tested, 202 participated in the second testing, 62 had died during the intervening 5 years, 32 were too ill to be retested (were in hospitals or had to remain in bed during the weeks of the testing), and 84 refused to be retested. A comparison of these four categories of subjects by age level is given in Table 1 and shows that the number of subjects retested decreased rather regularly with age, whereas the number of deceased

TABLE 1

FATE OF SUBJECTS FROM THE ORIGINAL SAMPLES AT THE TIME OF THE SECOND TESTING

	55-59	60-64	65-69	70-74	75+	Sum
Retested	51	42	44	34	25	202
Deceased	2	8	12	17	23	62
Too ill	4	2	6	11	9	32
Refused	19	18	14	14	19	84

subjects increased. The number of sick subjects increased irregularly. No systematic differences in the number of noncooperative subjects existed between the age levels.

Method

The following measures were used

1. Short forms of the Hamburg Wechsler Intelligence Test for Adults. These scales were administered to all subjects and the full test to a random subsample of 128 (see K. F. Riegel & R. M. Riegel, 1962; R. M. Riegel, 1960; R. M. Riegel & K. F. Riegel, 1959, 1962).

2. Five multiple-choice verbal tests (synonyms, antonyms, selections, classifications, analogies) as described by Riegel (1959, 1967). Though subjects were under no time stress, the duration of the test performances were recorded. Half of the items of the antonyms, selections, and classifications tests were mixed (m); the others were presented each in separate forms (s).

3. Four attitude and interest scales of the Likert type (rigidity, dogmatism, attitude toward life, interests) as described by Riegel and Riegel (1960).

4. A general questionnaire on the social and living conditions of the subjects, including inquiries on the following topics: education, financial status, health, physical activities, leisure-time activities, social activities, itemized activities, expressions of well-being, comparisons of situations, as well as many single items dealing with the conditions and habits of daily life (see Riegel et al., 1962).

RESULTS AND DISCUSSION

Developmental Trends of Subsamples

The main problems to be analyzed can be outlined in reference to Figures 1 and 2, which represent the average scores of the various subgroups on the scale of behavioral rigidity (see Riegel & Riegel, 1960). According to Figure 1, which includes the data from the first testing only, rigidity increased rather steadily with age. Subjects who were later retested were less rigid than the total group, but the rate of increase was about the same for both. Subjects who died during the following 5 years were much more rigid than both the total group and the subjects to be retested. Age differences were small for nonsurvivors. The scores of subjects who were

ill at the time of the second testing varied rather markedly between the age levels, but generally were far above those of the other groups. The same held for the subjects who later refused to be retested. However, this group did not deviate as much from the general trend as did the ill subjects.

Figure 2 compares the mean scores obtained at the second testing with total scores of the first test. Again the rigidity of the retested subjects increased with age. The trend of the means was closely parallel to that of the first testing, though the means were significantly higher ($p < .01$). Of greater importance, however, was the fact that the retest means were still below the averages of the total group at the first testing. This may be attributed to the absence of those subjects who died, became ill, or refused to be retested. If one reinstated these subjects by averaging their scores from the first testing with the second scores from the retested group, artificially complete samples would be created on the rather disadvantageous assumption that the scores of the subjects not retested would have remained unchanged during the intervening 5 years. The means of these artificial groups were comparable to those of the total group at the time of the first testing. As Figure 2 shows, they were above those of the first testing as well as above all other subgroups; that is, rigidity was higher during the later years than estimated in any of the other analyses. Of major concern in our discussion below will be the question of which of the various curves of Figure 2 would

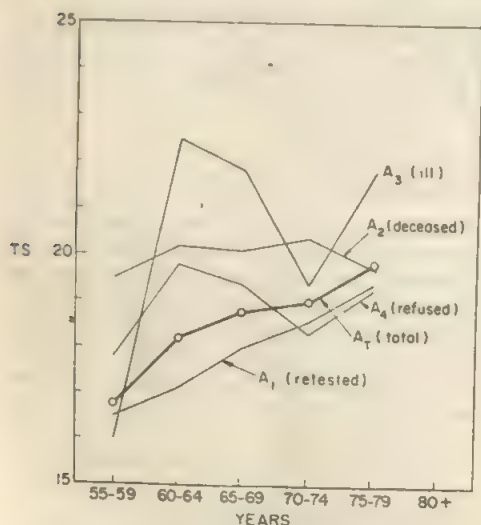


FIG. 1. Mean scores in behavioral rigidity at the first testing for five age levels and four subgroups of subjects.

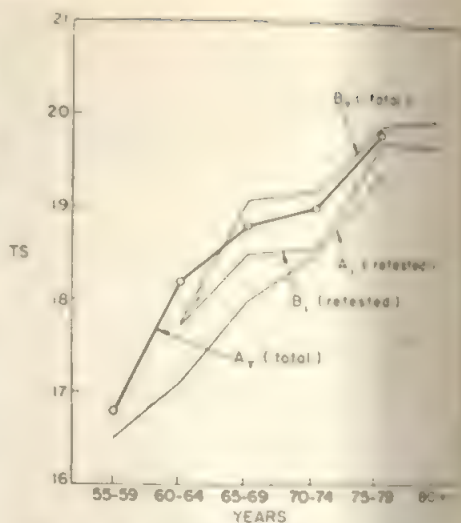


FIG. 2. Mean scores in behavioral rigidity at the five age levels of the total group and the group of retested subjects at the first (A) and second (B) testing.

represent most appropriately the developmental trend.

Dropout Rates

Since at the younger age levels some of the four subgroups did not include enough cases to allow for reliable estimates, the five consecutive age samples were pooled using the retirement age of 65 years as the cutting point. The significance of the retirement age for changes in behavior has been strongly suggested by the authors' as well as by earlier investigations. But aside from any such considerations, the age of 65 years subdivided the group of retested subjects in nearly equal sections of 99 and 133 subjects, respectively, and was selected primarily for this reason.

In the following analysis, the 43 variables tested are being regarded as a sample of measures not all of which are independent of one another but whose interdependencies—though slightly increasing with age—do not yield different correlation matrices at the two age levels. The variability in scores did not change markedly with age even though the dropout of subjects at the earlier age levels seemed to depend on systematic rather than random factors. These results justify the following comparisons in the numbers of significant differences between the subgroups.

For each variable the differences between the means of the four subgroups at the two remaining age levels have been tested for significance

TABLE 2

RELATIONSHIP OF DIFFERENCES IN MEAN SCORES BETWEEN AGE GROUPS
OF THE ORIGINAL SAMPLE*

Item	Age group 1 (N = 64)						Age group 2 (N = 63)					
	1-2	1-3	1-4	2-3	2-4	3-4	1-2	1-3	1-4	2-3	2-4	3-4
Age												
Intelligence												
Verbal												
Performance												
Verbal tests												
Synonyms												
Vocabulary												
Spelling												
Classifications												
Analogies												
Rigidity												
General												
Personal												
Dogmatism												
Anxiety												
Intolerance												
General												
Attitude toward life												
Pro retrospect												
Contemporary												
Interests												
Receptive activity												
Productive activity												
Physical activity												
Education												
Financial status												
Health												
Physical activity												
Social activity												
Leisure-time activity												
Itemized activity												
Well-Being												
Competitive situations												
2 signs	7	8	2	1	6	11	12	13	14	5	1	18

* The subgroups are denoted by the numbers and letters in the heading of the table: the following was 1 = retested subjects, 2 = decrease 1 subjects, 3 = subjects too ill, 4 = subjects refusing to be retested, A = same subjects, retested + (II) + refusing subjects, N = not retested subjects, decreased + 1 = refusing subjects. Thus, the difference 1-2 indicates the difference between subgroups 1 and 2, etc. If any of the differences was significant ($p < .05$) and positive a positive sign (+) was used; if negative, a negative sign (-) was given.

by analyses of variances. Subsequently, the most important t -test comparisons between the subgroups have been made and are explained in the footnote to Table 2. Signs are given whenever the differences are significant at or beyond the .05 level. In particular, positive signs indicate that the direction of differences in the first group of comparisons leaned toward greater intelligence, superiority in verbal abilities, more rigidity, more negativism in their attitude toward life, less interested, better education and financial support, poorer health, greater activity, feeling of well-being, and a favorable comparison of their present situation with the past.

By adding all the significant signs for each of the two age groups, it is evident that the subsamples deviated increasingly from one another. Below 65 years there were 35 significant differences, whereas for those over 65 there were 63 significant signs. This difference was particularly marked for the three intelligence measures, where the ratio of the signs for the two age groups was 3:12, as well as for the battery of verbal tests (ratio of 8:19) and the questionnaire scales (ratio of 5:13). An equal number of significant signs occurred for the two age groups on the various attitude scales (ratio of 19:19). Similar results were obtained by comparing the retested

and nonretested subjects (1 - N). Eleven of the 52 variables listed in Table 1 were significant for the younger, but 18 for the older age group.

In particular, both the deceased and the sick subjects differed on a rather large number of variables from the retested subjects. Below 65 years, seven and eight variables, respectively, differed significantly. Above 65 years, the corresponding figures were 12 and 13. For variables discriminating between noncooperative and retested subjects there was also a rather marked age difference. Below 65 years these two groups differed only on two scales (dogmatism and leisure-time activity). Above 65 years, however, 24 variables were significantly different. For both age groups, there were few significant differences between deceased and sick subjects. Only one variable was significant for subjects below 65, and five for subjects above 65 years of age.

Undoubtedly of greatest importance was the finding that in the younger group the deceased subjects differed from all the survivors (including those retested or not retested) on six variables, whereas there was only one significant difference at the higher age level. This finding encouraged a search for systematic differences that might have allowed the prediction of the subjects' death on the basis of their sociopsychological functioning at the time of the first testing.

Prediction of Death

Multiple-regression predictions of death were calculated on an IBM 7090 computer. Age and sex were excluded from the predictions since both of them—being known as good predictors—could have covered up the more interesting psychological variables, even though their elimination would necessarily reduce the overall degree of correlation.

The predictions were made by successively reducing the F level at which variables would enter into the equations. The F levels were set at .05 and .10, but for each variable the F values were also empirically determined. This was necessary because under the more lenient conditions of $p = .10$ the F values for some of the variables already entered could have changed, depending on their correlations with the newly entering measures. Indeed, in some cases, high-ranking predictor variables have been dropped altogether from the equations. Generally, in a new field of investigation it seems reasonable to go somewhat beyond the conventional level of $p = .05$ because additional variables will pick up unaccounted portions of the variance. Even though they will increase the likelihood of Type 2 errors, it seems more appropriate to retain hypotheses initially, which after sufficient further research turn out to be false, rather than to

TABLE 3
RANKS AND F LEVELS FOR THE PREDICTION OF DEATH
AT TWO AGE LEVELS

Variable	55 to 64		65 and over	
	Rank	F	Rank	F
Personal rigidity	11	4.51*		
General rigidity	10	3.10*		
General dogmatism	2	6.91	2	8.80
Productive activities	4	5.48	4	6.75
Classifications m			5	5.70
Antonyms T	3	9.37		
Classifications T	5	4.42		
Total testing time m	7	4.48		
Financial status	9	3.01*		
Health	1	11.94		
Physical activity			1	8.80
Free time	8	3.16*		
Yrs. married			5	4.80
No. acquaintances	12	3.57*		
Itemized activities	6	3.91		
Widow(er) or not	13	2.76*		

* $p = .10$.

reject hypotheses too early that may turn out to be true.

As shown in Table 3, 13 variables entered into the equation for the 55- to 64-year-old subjects at an F level of .10, but only five for the subjects above 65 years. Moreover, one of these five variables correlated highly with age, namely, the number of years married, and thus must have been excluded for the same reason for which age has been eliminated. Only two of the variables were also predictors for the younger age group, namely, general dogmatism and productive activities (interests). The remaining variables were different for both groups. Even though the best single predictors were health for the younger, and the amount of physical activity for the older group, both the attitude and the intellectual measures entered strongly into the regression equations. Because the various intercorrelations were taken into account, the variables selected and shown in Table 3 were not necessarily identical with those significantly different in means between deceased and live subjects (see Table 2).

The point-biserial multiple correlations at the F level of .05 were .47 for the seven variables selected into the equation of the younger, and .31 for the three variables of the older subjects. When the F level was lowered to .10, the multiple correlations were raised to .60 and .55, respectively. In further lowering the F level to .50, the correlations could have been increased to .68 and .46, respectively, but at the same

the errors of estimate would have grown considerably so that this gain would not have been beneficial.

CONCLUSIONS

The present analysis was based on five cross-sectional samples of consecutive age groups above 45 years. Five years after the first testing, attempts were made to retest all subjects. However, success was not possible in these attempts because, in numbers increasing with age, some subjects had died or had become sick. Others refused to be retested, but their numbers seemed to be independent of age. If the biological determinants of death and disease interact with psychological or sociological factors—as has been shown in the present study—a number of far-reaching conclusions can be drawn.

First, a concept of development as represented by curves of growth and decline is questionable, regardless of whether these curves are based on longitudinal or cross-sectional data.

Development as reflected by trends of average test or rating scores would be a meaningful concept only if the dropout of subjects in consecutive age groups is strictly a random process. However, biologically weak subjects will die earlier or will become increasingly unable to participate in the testings. A particular occupational, financial, educational and/or social status may increase the risks and may decrease the chances of survival. According to the present study, groups of dropouts (particularly the younger ones) are also psychologically different in their abilities and attitudes.

Age samples drawn for cross-sectional studies or followed up in longitudinal studies become increasingly biased, the further one moves upward in the age scale, and thus the generalized trend will be confounded by the increasing degree of sample bias.

Second, older reports on the decline of intelligence have been received with shock or suspicion. Relief was felt when some longitudinal studies (Bayley & Oden, 1955; Owens, 1953) reported higher stabilities than originally observed, even though these studies were restricted to subjects with superior capacities. According to the present findings and aside from other psychological factors, less-able persons die earlier or are more likely to become seriously ill. Thus, the samples are increasingly loaded with highly able persons. If it had been possible to retest all of the subjects originally involved in the study, the decline would have been more marked than either cross-sectional or longitudinal studies have revealed.

A third conclusion can be added concerning the determinants of death. Even though the

amount of disease suffered by the subjects is a function of numerous problems of disease with more knowledge of the disease than at the time age began. This seems to contradict the general belief that intelligence declines with increasing age, and to the larger differences in average age scores between the groups of subjects who were either sick, confined to a hospital, or had died. The differences in samples of populations indicate that at lower age levels dropouts can be described as a subgroup which differs systematically from the rest in almost the same way as subjects of higher age. Attacks may arise from various causes. At the higher age levels, however, death seems to occur at random, and the trend of intelligence becomes less valid. Using data for the present study, it is impossible to trace the persons who passed the degrees of aging rather than age to disease which cause their death.

REFERENCES

- BAYLEY, N. & ODEN, M. H. The stability of intellectual ability in adult males. *Journal of Gerontology*, 1955, 10, 30-37.
- BELL, A. & ZIMM, J. P. The effect of age on intellectual performance in normal declining individuals. *Journal of Gerontology*, 1955, 15, 101-105.
- BERKOWITZ, B. & GREEN, R. F. Changes in health with age. I. Longitudinal study in Western Baltimore. *Journal of General Psychology*, 1955, 103, 3-11.
- DIAMOND, A. Relationships between biological, psychological and environmental studies in adult man. *Physiology and Psychology: Human Development*, 1955, 8, 16-22.
- ENDERBURY, C. The WAIS performance of the aged: A recent evaluation. *Journal of Gerontology*, 1955, 18, 169-171.
- FARLEY, A., KALLMAN, F. J., JARVIE, L. & JARVIE, L. F. Longevity and intellectual variation in a senescent twin population. *Journal of Gerontology*, 1955, 15, 305-307.
- FOSTER, J. C. & TAYLOR, G. A. The application of mental tests to persons over 60. *Journal of Applied Psychology*, 1932, 4, 40-45.
- JARVIE, L. F. & FARLEY, A. Intellectual changes and survival in the aged. *Journal of Gerontology*, 1955, 18, 174-176.
- JARVIE, L. F., KALLMAN, F. J. & FARLEY, A. Intellectual changes in aged twins. *Journal of Gerontology*, 1962, 17, 289-294.
- JARVIE, L. F., KALLMAN, F. J., FARLEY, A. & KLABER, M. M. Changing intellectual functions in senescent twins. *Acta Genetica et Statistica Medica*, 1957, 7, 421-430.
- JONES, H. E. & CONRAD, H. S. The growth and decline of intelligence: A study of a homogeneous group between the ages ten and sixty. *Genetic Psychology Monographs*, 1935, 13, 225-268.
- JONES, H. E., CONRAD, H. S. & HORN, A. Psychological studies of motion pictures. II. Observation and recall as a function of age. *University of*

- California Publications in Psychology, 1928, 3, 225-243.
- KAPLAN, O. J. Mental decline in older morons. *American Journal of Mental Deficiency*, 1943, 47, 277-285.
- KAPLAN, O. J. *Mental disorders in later life*. Stanford: Stanford University Press, 1956.
- KAPLAN, O. J., RUMBAUGH, D. M., MITCHELL, D. C., & THOMAS, E. D. Effects of level of surviving abilities, time of day, and test-retest upon psychological performance in seniles. *Journal of Gerontology*, 1963, 18, 55-59.
- KLEEMEIER, R. W. Intellectual changes in the senium. In *Proceedings, Social Statistics Section, American Statistical Association*. Washington, D. C.: American Statistical Association, 1962. Pp. 290-295.
- MILES, W. R. Age and human ability. *Psychological Review*, 1933, 40, 99-123.
- OWENS, W. A., JR. Age and mental abilities: A longitudinal study. *Genetic Psychology Monographs*, 1953, 48, 3-54.
- RIEGEL, K. F. A study on verbal achievements of older persons. *Journal of Gerontology*, 1959, 14, 453-456.
- RIEGEL, K. F. Changes in psycholinguistic performance with age. In G. A. Talland (Ed.), *Human behavior and aging: Recent advances in research and theory*. New York: Academic Press, 1967, in press.
- RIEGEL, K. F., & RIEGEL, R. M. A study on changes of attitudes and interests during later years of life. *Vita Humana*, 1960, 3, 177-206.
- RIEGEL, K. F., & RIEGEL, R. M. Analysis of differences in test performance and item difficulty between young and old adults. *Journal of Gerontology*, 1962, 17, 97-105.
- RIEGEL, K. F., RIEGEL, R. M., & MEYER, G. Sociopsychological factors of aging: A cohort-sequential analysis. *Human Development*, 1967, 10, in press.
- RIEGEL, K. F., RIEGEL, R. M., & SKIBA, G. Untersuchung der Lebensbedingungen, Gewohnheiten und Anpassung älterer Menschen in Norddeutschland. *Vita Humana*, 1962, 5, 204-247.
- RIEGEL, R. M. Faktorenanalysen des Hamburg-Wechsler-Intelligenztests für Erwachsene (HAWIE) für die Altersstufen 20-34, 35-49, 50-64 und 65 Jahre und älter. *Diagnostica*, 1960, 6, 41-66.
- RIEGEL, R. M., & RIEGEL, K. F. Standardisierung des Hamburg-Wechsler-Intelligenztests für Erwachsene (HAWIE) für die Altersstufen über 50 Jahre. *Diagnostica*, 1959, 5, 97-128.
- RIEGEL, R. M., & RIEGEL, K. F. A comparison and reinterpretation of factorial structures of the W-B, the WAIS, and the HAWIE on aged persons. *Journal of Consulting Psychology*, 1962, 26, 31-37.
- ROSE, C. H. Representatives of volunteer subjects in a longitudinal aging study. *Human Development*, 1965, 8, 152-156.
- SCHAE, K. W. Cross-sectional methods in the study of psychological aspects of aging. *Journal of Gerontology*, 1959, 14, 208-215.
- SCHAE, K. W. A general model for the study of developmental problems. *Psychological Bulletin*, 1965, 64, 92-108.
- SCHAE, K. W., & STROTHER, C. R. The effect of time and cohort differences on the interpretation of age changes in cognitive behavior. *American Psychologist*, 1964, 19, 546.
- SUSSMAN, M. B. Use of longitudinal designs in studies of long-term illness, some advantages and limitations. *Gerontologist*, 1964, 4, 25-29.

(Received January 7, 1966)

Journal of Personality and Social Psychology
1967, Vol. 5, No. 3, 348-350

FACTOR STRUCTURE OF DUNCAN'S PERSONALITY INTEGRATION SCALE

LOGAN WRIGHT¹

Purdue University

Duncan's Personality Integration Reputation Test was administered to 3 groups of Ss consisting, respectively, of 43 sorority members, 92 fraternity members, and 129 sorority members at a large midwestern university. These data were analyzed in order to answer the following questions: (a) What is the factor structure of this scale; (b) is the structure reliable; and (c) is the factor structure related to sex differences? Obtained data indicated that the scale is comprised of an extremely potent 1st factor which accounted for approximately 90% of the total variance associated with a given matrix. In addition, replications showed the factor structure to be reliable and also not susceptible to sex differences.

In the last 10 years, positive mental health and effective behavior have commanded increased attention within the areas of clinical psychology

¹Now at the University of Oklahoma Medical Center.

and personality theory (Barron, 1955; Jahoda, 1958; Maslow, 1954; McQuitty, 1954; Seeman, 1959; Shoben, 1957; Smith, 1959; White, 1959; Wishner, 1955). To date, a six-item sociometric device constructed by Duncan (1966) is the most

extensively developed instrument designed specifically to assess this type of behavior. Research in positive mental health would be facilitated greatly by the availability of objective criterion measures whose validity and factor structures are known. Previous studies (Duncan, 1966; Seeman, 1966; Wright, 1966) have strongly supported the validity of Duncan's scale. The present study was designed to answer three questions: (a) What is the factor structure of the Duncan scale; (b) is this structure reliable; and (c) is the factor structure related to sex differences?

Initially, Duncan's Personality Integration Reputation Test was administered to 43 sorority members at a large midwestern university. The item intercorrelation matrix generated from these data was factored according to a principal components solution. R^2 was employed to estimate communalities. The unrotated factor matrix revealed a first factor which accounted for 84% of the total matrix variance. The magnitude of the loading for each item of the scale on this factor exceeded .40, which had been established as an arbitrary index of significance (see Table 1). No other factor could account for 7% of the total variance.

To replicate this finding, and to assess the influence of sex differences, the scale was next administered to a group of 92 single, white undergraduate male fraternity members and to a group of 129 single, white female undergraduate sorority members at a large midwestern university. The data obtained for males, for females, and for both groups combined were factor analyzed by means of the same principal components solution that was utilized in the initial analysis. In each case, males, females, and both groups combined, a first factor was extracted that accounted for 88%, 91%, and 91% of the total matrix variance, respectively. In all analyses, the magnitude of the loading for each item of the scale on the first factor exceeded .58 (see Table 1). In no case did another factor account for more than 4% of the total variance associated with a given matrix. Since this potent first factor was obtained without exception, no factor matrix rotation was obtained.

The manner in which Duncan's scale was derived should lend some light to the interpretation of the obtained first factor. The items comprising the instrument were derived from a review of literature concerning positive mental health by Jahoda (1958). Jahoda's review yielded six categories which theorists and researchers have suggested to be related to positive mental health. Duncan then constructed an item for each of the six component areas of positive mental health. Thus, the derived first factor appropriately might be regarded as a measure of positive mental

TABLE 1

FACTOR LOADINGS OF THE SIX DUNCAN SCALE ITEMS ON THE FIRST FACTOR DERIVED IN FOUR ANALYSES

Duncan scale items	Loadings obtained with 43 females	Loadings obtained with 92 males	Loadings obtained with 129 females	Loadings obtained with group of 12 males and 129 females
1	.66	.79	.79	.80
2	.59	.67	.81	.74
3	.60	.72	.80	.77
4	.58	.63	.72	.68
5	.54	.58	.69	.65
6	.46	.58	.73	.66
% of variance explained by 1st factor	84%	88%	91%	91%

health or effectiveness in an omnibus or inclusive sense. However, since the scale is a sociometric measure in which subjects' scores are based on the ratings of others, the factor must be viewed as a measure of perceived health or effectiveness. The derived factor was therefore regarded as "effectiveness stimulus value."

The data obtained in this study support the notion that positive mental health, as defined by workers in the field and as measured by Duncan's scale, can be regarded as a unitary or general factor. In addition, these data indicate that the factor structure of the scales may be replicated, and that the structure is not influenced by sex differences.

REFERENCES

- BARRON, F. Toward a positive definition of psychological health. Paper read at American Psychological Association, San Francisco, September 1955.
- DUNCAN, C. B. A reputation test of personality integration. *Journal of Personality and Social Psychology*, 1966, 3, 516-524.
- JAHODA, M. *Current concepts of positive mental health*. New York: Basic Books, 1958.
- MASLOW, A. H. *Motivation and personality*. New York: Harpers, 1954.
- MCQUITT, L. L. Theories and methods in some objective assessments of psychological well-being. *Psychological Monographs*, 1954, 68(14, Whole No. 385).
- SEEMAN, J. Toward a concept of personality integration. *American Psychologist*, 1959, 14, 633-637.
- SEEMAN, J. Personality integration in college women. *Journal of Personality and Social Psychology*, 1966, 4, 91-93.
- SHOBEN, E. J., JR. Toward a concept of the normal personality. *American Psychologist*, 1957, 12, 183-189.

SMITH, M. B. Research strategies toward a conception of positive mental health. *American Psychologist*, 1959, 14, 673-681.

WHITE, R. W. Motivation reconsidered: The concept of competence. *Psychological Review*, 1959, 66, 297-333.

WISHNER, J. The concept of efficiency in psychological health and in psychopathology. *Psychological Review*, 1955, 62, 69-80.

WRIGHT, L. Construct validity of Duncan's personality integration scale. *Educational and Psychological Measurement*, 1966, 26, 471-478.

(Received February 10, 1966)

Journal of Personality and Social Psychology
1967, Vol. 5, No. 3, 350-352

HONESTY OF SUBJECTS AND BIRTH ORDER¹

PAUL LANE WUEBBEN

Department of Sociology, University of California, Santa Barbara

The common assumption that Ss who promise not to talk to others about the research in which they are participating do not talk was tested. Post hoc analyses of data from 113 male experimental Ss showed that 72 of them broke their promise and talked to others. Firstborn and only-child Ss talked to fewer others than did later-born Ss.

Social psychologists have recently become increasingly concerned with the methodological problems of laboratory experiments with human subjects. Several investigators, among them Rosenthal (e.g., Rosenthal, 1964; Rosenthal, Kohn, Greenfield, & Carota, 1966) and Sarason (Sarason, 1962; Sarason & Minard, 1963; Winkel & Sarason, 1964), have reported studies which indicate that the experimenter himself—his expectancies, desires, and personal attributes—may be an important, though unintended, source of variance in experimental findings. Riecken (1962) and Orne (Orne, 1962; Orne & Evans, 1965; Orne & Scheibe, 1964) have discussed and demonstrated both random and systematic biases associated with subjects' perceptions of experiments and their motivations to do what is "expected" of them. Other students have suggested that experimental manipulations often fail in their intent because investigators do not adequately control for subjects' perceptions and definitions of the experimental situation and of the manipulations themselves (Chapanis & Chapanis, 1964; Elms & Janis, 1965; Janis & Gilmore, 1965; Rosenberg, 1965). The present paper identifies another methodological problem of laboratory experiments with human subjects.

¹Data were collected while the author was engaged in dissertation research in the Department of Sociology, University of Wisconsin. The financial assistance, during this period, of the Vilas Fellowship Fund of the University of Wisconsin is gratefully acknowledged. Analyses were facilitated by a grant from the Committee on Research, University

Many, if not most, laboratory experiments with human subjects depend for their effectiveness upon the postexperimental cooperation of earlier run subjects. Thus, to insure that later run subjects will be uninformed about the manipulations, deceptions, and measurements involved in his particular experiment, the researcher usually elicits from all subjects a promise not to talk to others about the experiment until it has been completed. This paper examines the extent to which such promises are kept.

METHOD

Subjects

Subjects were 113 male students who volunteered in their elementary sociology courses at the University of Wisconsin, to participate in a "university-wide" research study. Subjects who participated were given extra credit in their courses. Approximately 90% of those asked did volunteer.

Data Collection

Data were collected in an experimental study which was designed to test hypotheses other than those to which this paper is addressed. Full explanation of the experimental procedure can be obtained elsewhere (Wuebben, 1966). However, it should be mentioned that the experimental context within which these data were secured involved the inducing of fear in subjects; the experiment was designed primarily to explore the relationship between fear and acceptance of a fear-reducing,

of California, Santa Barbara Division. The author wishes to thank Bob Terry and Don Zimmerman for their critical comments.

persuasive communication. No statistically significant relationships were found between the dependent variables discussed in this paper and the independent variables of major interest in the experiment (level of severity of threat and level of probability of threat).

Procedure

Each subject attended two experimental sessions; the second session was run 1 week after the first. All subjects were run through each session within a 27-hour period.

In the first experimental session, subjects were run in groups of between 14 and 21 in size. When all subjects had arrived at the meeting place, a lecture room, they were greeted by the "researcher" (in reality an actor) who identified himself as "Dr. Goodrich" of the Department of Preventive Medicine and Psychology in the Medical School of the University of Wisconsin. Dr. Goodrich explained that his university-wide research project had several related purposes. One purpose of the research was held to be the collection of information about how students reacted to the possibility of contracting a certain disease which would be described to them. The subjects were told that the information they provided would be used, later, in organizing a nationwide information and education campaign about the disease.²

The rationale given subjects for answering questions gave emphasis to the importance of their not talking to others about the study. At the end of the first experimental session, Dr. Goodrich made the following plea to the subjects:

O.K., I guess that about takes care of everything for tonight. Oh, just one thing before you go. Other people will be going through this orientation session. If we are going to get any useful results on those questions we asked you tonight, it is necessary that these other people get their information from *us*. In other words, we don't want someone in the next session having any prior information or misinformation about the disease we discussed or about this study. All this means is that we need the cooperation of each and every one of you in a pledge not to talk about this study to *anyone* until it is announced in class that it has officially ended. Please do not talk to anyone, even if you know they're not in the study. You never know who might hear about things indirectly. So I'll ask each of you to promise not to mention to anyone what this study is about. Will that be O.K.?

Dr. Goodrich nodded his head and indicated that he wanted the subjects to nod their heads also if they agreed not to talk about the study. Dr. Goodrich concluded his talk only after he had observed each subject indicate that he would not talk.

² Data not reported here showed that subjects did not question either the credibility of Dr. Goodrich or the rationale of the study.

In the second experimental session, subjects spent some 40 minutes answering a battery of questions. Subjects were not told about the true nature of the study in which they had participated until after they had answered all questions. One of those questions was designed to measure both whether or not subjects had talked to others about the study, and, if they had talked, the number of others to whom they had spoken. Subjects were asked the following question:

As you recall, you were asked not to mention to anyone what this study was about. Quite frankly, we knew it would be natural for many people to talk about the study; however, the usefulness of the results was not diminished even though a few subjects did know what the research concerned. If you did talk to others, it would help us to know how many. No one will be penalized or rewarded in any way for having talked. I ESTIMATE THAT I TALKED TO
..... PEOPLE ABOUT THIS RESEARCH.

RESULTS AND DISCUSSION

Both the author and the actor who played Dr. Goodrich observed subjects' responses to Dr. Goodrich's request that they promise not to talk. The observers agreed that all subjects indicated by nodding their heads that they would not talk to others about the study.

Of the 113 subjects who participated in the experiment and publicly promised not to talk, 72 or 64% admitted that they had talked to at least one other person about the experiment during the week between the first and second experimental sessions; 41 subjects or 36% said that they had talked to no one. Further, most subjects who did talk talked to more than one person about the study in which they were participating. Thirty-three subjects indicated that they had talked to only one person; 39 subjects admitted that they had talked to more than one. The range in the number of people to whom subjects talked was from 1 to 10. Those who talked talked to a mean of 2.24 people about the research. The common assumption that subjects who promise not to talk to others about an experiment do not talk is not supported by these data!

An analysis of variance (corrected for unequal *N*s) showed that firstborn and only-child subjects talked to fewer persons ($M = .93$) than did later-born subjects ($M = 1.72$, $F = 4.76$, $df = 1/111$, $p < .05$). This finding is congruent with the results of several studies (e.g., Arrowood & Amoroso, 1965; Becker, Lerner, & Carroll, 1964, 1966; Schachter, 1964; Staples & Walters, 1961) which have shown that firstborn and only-child persons are more dependent, more

conforming, and more responsive to normative pressures than are later-born persons.

The methodological implications of present findings are, potentially, of great importance. Social psychologists have only recently come to realize how social interaction between subjects and experimenters *during* an experiment may both bias findings and make difficult their interpretation. The present study points to another methodological complication—even though subjects may promise not to talk to others about a study in which they are participating, they do talk. As a consequence, students who come to later runs of an experiment may not only share general societal norms about “psychological studies,” but they also may be in possession of “expert” detailed knowledge about the particular study in which they are going to participate.³

Present findings also throw open to doubt the common assumption that inexperienced beginning students are necessarily naive about laboratory experimentation. However, the extent to which student communication about any given experiment might bias the results of other experimental studies cannot be determined from available data. The topic seems important enough to warrant careful empirical investigation of the relevant student norms.

³ Rosenthal et al. (1966) found that subjects run in the second session of their experiment were more aware of the fact that they had undergone a verbal conditioning procedure than were first-run subjects. The authors suggested that communication from Session I subjects to Session II subjects may have been responsible.

REFERENCES

- ARROWOOD, A. J., & AMOROSO, D. M. Social comparison and ordinal position. *Journal of Personality and Social Psychology*, 1965, 2, 101-104.
- BECKER, S. W., LERNER, M. J., & CARROLL, J. Conformity as a function of birth order, payoff, and type of group pressure. *Journal of Abnormal and Social Psychology*, 1964, 69, 318-323.
- BECKER, S. W., LERNER, M. J., & CARROLL, J. Conformity as a function of birth order and type of group pressure: A verification. *Journal of Personality and Social Psychology*, 1966, 3, 242-244.
- CHAPANIS, P., & CHAPANIS, A. Cognitive dissonance: Five years later. *Psychological Bulletin*, 1964, 61, 1-22.
- ELMS, A. C., & JANIS, I. L. Counter-norm attitudes induced by consonant versus dissonant conditions of role playing. *Journal of Experimental Research in Personality*, 1965, 1, 50-60.
- JANIS, I., & GILMORE, J. The influence of incentive conditions on the success of role playing in modifying attitudes. *Journal of Personality and Social Psychology*, 1965, 1, 17-27.
- ORNE, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 1962, 17, 776-783.
- ORNE, M. T., & EVANS, F. J. Social control in the psychological experiment: Antisocial behavior and hypnosis. *Journal of Personality and Social Psychology*, 1965, 1, 189-200.
- ORNE, M. T., & SCHEIBE, K. E. The contribution of nondeprivation factors in the production of sensory deprivation effects: The psychology of the “panic button.” *Journal of Abnormal and Social Psychology*, 1964, 68, 3-12.
- RIECKEN, H. W. A program for research on experiments in social psychology. In N. F. Washburne (Ed.), *Decisions, values and groups*. Vol. 2. New York: Pergamon Press, 1962. Pp. 25-41.
- ROSENBERG, M. When dissonance fails: On eliminating evaluation apprehension from attitude measurement. *Journal of Personality and Social Psychology*, 1965, 1, 28-42.
- ROSENTHAL, R. The effect of the experimenter on the results of psychological research. In B. A. Maher (Ed.), *Progress in experimental personality research*. New York: Academic Press, 1964. Pp. 79-114.
- ROSENTHAL, R., KOHN, P., GREENFIELD, P. M., & CAROTA, N. Data desirability, experimenter expectancy, and the results of psychological research. *Journal of Personality and Social Psychology*, 1966, 3, 20-27.
- SARASON, I. G. Individual differences, situational variables, and personality research. *Journal of Abnormal and Social Psychology*, 1962, 65, 376-380.
- SARASON, I. G., & MINARD, J. M. Interrelationships among subjects, experimenter, and situational variables. *Journal of Abnormal and Social Psychology*, 1963, 67, 87-91.
- SCHACHTER, S. Birth order and sociometric choice. *Journal of Abnormal and Social Psychology*, 1964, 68, 453-456.
- STAPLES, F. R., & WALTERS, R. H. Anxiety, birth order, and susceptibility to social influence. *Journal of Abnormal and Social Psychology*, 1961, 62, 716-719.
- WINKEL, G. H., & SARASON, I. G. Subjects, experimenter, and situational variables in research on anxiety. *Journal of Abnormal and Social Psychology*, 1964, 68, 601-608.
- WUEBBEN, P. L. Experimental analysis of threat-based persuasive appeals. Unpublished doctoral dissertation, University of Wisconsin, 1966.

(Received April 14, 1966)

INFORMATIONAL PROPERTIES OF VERBAL AND NONVERBAL EVENTS¹

ROBERT B. CAIRNS

Indiana University

The principal question explored in this experiment was whether the reinforcement effectiveness of an established verbal reward could be enhanced by structuring its informational content through instructions. 40 4th grade children were administered a modified version of the Wisconsin Card Sorting Test. In a 2×2 factorial design, the outcome events used differed in terms of previously acquired informational properties ("right" and a buzzer of moderate loudness), and in terms of experimentally induced informational content (structured and unstructured). The results indicated a strong experimentally induced information effect ($p < .001$) for both the verbal and nonverbal events, with the latter showing the greatest relative gain in effectiveness.

This paper is concerned with some determinants of the effectiveness of approval-assert words in the control of children's behavior. Recent interest in a child's response to verbal rewards (see reviews by Stevenson, 1965; Walters & Parke, 1964) has been supported by the assumption that these events are immediately relevant to one or more conditioned motives of the child. Words such as "right," "fine," and "good" thus have been assumed to be effective as reinforcement events because they reduce an appetitive (e.g., need for approval, dependency, social drive) or an aversive (e.g., anxiety) drive. Following the lead of Gewirtz and Baer (1958), several investigators have attempted to demonstrate experimentally that a given motive is critical in controlling verbal event effectiveness.

Current reviews of the area suggest, however, that the work has failed in its original aim of clarifying the conceptual status of secondary motives. At least two reasons can be offered for this inconclusive state of affairs. First, as Parton and Ross (1965) have shown, the area has been plagued by a variety of methodological pitfalls. Among other problems, appropriate experimental and statistical controls have been regularly omitted. Second, the differential effects of experimental treatments upon verbal reward effectiveness apparently are not very strong. Even among those reports that have claimed statistical significance, the proportion of the total variance that can be attributed to the main experimental effect is often surprisingly small.²

In view of the inconclusiveness of the empiri-

cal results, it is difficult to justify the hegemony, in the personality and developmental literature, of motivational accounts of verbal event effectiveness. There has been, in fact, a recent trend to give somewhat more attention to the cue properties of verbal reinforcement events (e.g., Spence, 1966; Walters & Parke, 1964). Relevant to this emphasis upon the signal properties of the word (as opposed to the motivational structure of the child) is the information hypothesis of Postman and Sassenrath (1961). These authors have proposed, essentially, that verbal events are effective primarily by virtue of the information which they impart to the learner (cf. Egger & Miller, 1962).

This paper extends the analysis of the cue properties of verbal and nonverbal events. The particular hypothesis tested was that the reinforcement effectiveness of both classes of events could be enhanced by a brief instructional set which reduced their ambiguity. Of the two types of outcome events studied, one (the word "right") presumably had some previously established informational properties; the other, a buzzer of moderate loudness, presumably had minimal previously established informational properties. The effectiveness of both outcomes was expected to be augmented by an informational set, with the nonverbal event showing the greatest relative gain over a noninstructed control group.

METHOD

Subjects

Forty fourth-grade children, 21 boys and 19 girls, served as subjects. The mean age was 9.4 years; the mean IQ (Stanford-Binet, Form LM, or Wechsler Intelligence Scale for Children) was 109.4. Children in each sex group were assigned at random to the four experimental conditions.

have had in replicating the effects of isolation and frustration (R. B. Cairns, "Antecedents of Social Reinforcer Effectiveness," Progress Report No. M-4373, United States Public Health Service, 1962).

¹ The author expresses his thanks to L. F. Read, superintendent, Bloomington Metropolitan Schools, and to J. Fleener, principal of Hunter School, for their help in making subjects and facilities available. This work was supported in part by funds from the National Institute of Mental Health, United States Public Health Service (Grant 07144-01).

² The weakness of the primary phenomenon may account for the difficulty that some investigators

Experimental Groups

Subjects were administered a modified version of the Wisconsin Card-Sorting Test (WCST—Grant & Berg, 1948). The experimental conditions were concerned with the nature of the outcome event following a "correct" response. Four different outcomes occurred:

1. Nonverbal-Unstructured: buzzer onset, without prior instructions that it signaled a correct response.
2. Nonverbal-Structured: buzzer onset, with prior instructions that it would occur only when a correct response was made.
3. Verbal-Unstructured: the experimenter said, "right," without prior instructions that this event signaled a correct response.
4. Verbal-Structured: the experimenter said, "right," with prior instructions that this event would occur only when a correct response was made.

The design was thus basically a 2×2 factorial model, with two levels of stimulus type (verbal and nonverbal) and two levels of informational set (structured and unstructured). The four groups did not differ in terms of age or intelligence. The two sexes were equally represented in each group, except for the Nonverbal-Unstructured group which included six males and four females.

Apparatus and Materials

Two standard WCST decks were combined to provide 70 response cards. Each card had one or more identical objects of a single color. Cards differed in terms of the shape of the objects (stars, circles, triangles, or crosses), color of the objects (green, blue, red, or yellow), and number of identical objects (1, 2, 3, or 4). In addition, four stimulus cards were affixed to a response board. Moving from the subject's left to his right, the cards on the response board had two green stars, four blue circles, one red triangle, and three yellow crosses. The board itself was a rectangular piece of heavy white cardboard (12×18 inches), with four double compartments spaced equally across its width. The stimulus cards were placed in the upper halves of the compartments; the lower halves were empty.

The other piece of equipment was a modified 12-volt buzzer which produced a noise of moderate intensity (approximately 65 decibels at the subject's ear). It was activated manually by a foot pedal.

Procedure

Prior to the experiment, the experimenter was introduced to the children of two fourth-grade classrooms by their teachers. For testing purposes, subjects were brought individually from their classrooms. Upon reaching the testing room, the experimenter said:

Please sit down [motions towards chair] and let me show you what you'll be working on. You are to take each of these cards and put it here [pointing], into one of the boxes beneath the four cards. Put the cards where you think they belong.

For those subjects in the "structured" conditions, the additional statement was made: "When you hear me say *right* (or this buzzer), you will know that you have put a card where it belongs." The

words, "this buzzer," were followed by the presentation of the nonverbal stimulus; the word *right* was emphasized by the experimenter.

For all subjects, the experimenter then said: "Do you have any questions? [pause] O.K., you may begin now." Questions were dealt with by the experimenter's repeating part or all of the instructions. Within these limits, care was taken by the experimenter to insure that the subject understood the task requirements.

A small table separated the experimenter and the subject. The response board was located on the table, immediately in front of the subject. The cards to be sorted were placed beside the response board.

Testing was divided into two phases, acquisition and extinction. During the reinforcement phase (the first blocks of 10 trials), the experimenter said "right" or activated the buzzer for approximately 50 msec immediately after the subject made a correct choice (i.e., matched the color of the response card selected on that trial to the color of the stimulus card). The response card was then removed by the experimenter. Following an incorrect choice, the experimenter simply removed the response card. During the final 20 trials (51–70), the extinction phase, the verbal, and nonverbal outcome events were omitted. The transition between the two test phases was not accompanied by a break in testing; once the test had begun, subjects sorted cards continuously for 70 trials. Subjects were not told, either before the test or during it, that the outcome events would be omitted after Trial 50.

Prior to the experiment, the cards were randomized through repeated shuffling. The cards were presented to all subjects in the same fixed random order.

RESULTS

Acquisition Performance

The number of correct responses produced by the four groups is shown in Figure 1. An analysis of variance of the acquisition data was carried out, with stimulus type (verbal-nonverbal), informational set (structured-unstructured), and trial blocks (1–5) as the main factors. In this analysis, the main effect of informational set was highly significant, $F(1, 36) = 57.41, p < .001$. The only other statistically reliable effects were those of trial blocks, $F(4, 144) = 38.63, p < .001$, and the Trial Blocks \times Information interaction, $F(4, 144) = 11.13, p < .001$. The latter results indicated that the total group improved in accuracy as a function of trials, with the "informed" conditions showing the greatest improvement. None of the other effects (i.e., stimulus type or any interaction involving stimulus type) was statistically significant in this analysis.

One important feature of the data was obscured by grouping the trials into blocks of 10, namely, the very rapid learning of subjects in the two structured conditions. As shown in Figure 2, all but 1 of the 20 subjects in the structured conditions responded in a fashion consistent with a "color" strategy on Trial 12 (and thereafter).

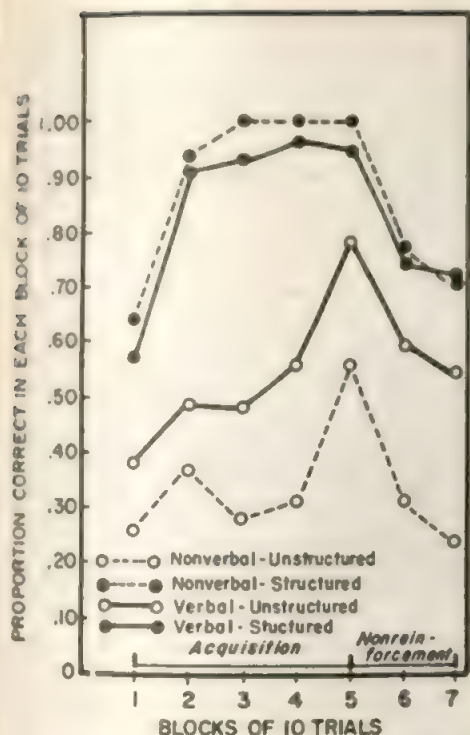


FIG. 1. The mean performance of the four experimental groups during acquisition (Blocks 1-5) and nonreinforcement (Blocks 6-7).

Conversely, only 3 of the 20 subjects in the unstructured conditions sorted on the basis of color on Trial 12 ($\chi^2_c = 22.73$, $df = 1$, $p < .001$).

A comment on the irregularity of the curves in Figure 2 is in order. The upswing in the proportion of correct responses at Trials 2 and 8 for all groups was primarily due to an overlap between the initially dominant strategy of shape and the criterion strategy of color. Analysis of the Trial 1 choices indicates that a "shape" strategy was initially preferred; that is, it was followed by 31 of the 40 subjects (78%). Furthermore, on 14 of the 50 acquisition trials, the two response strategies overlapped; that is, the response card had green stars, blue circles, red triangles, or yellow crosses. Subjects who persisted in sorting according to shape were therefore credited with having made a correct "color" choice on 2, 3, 2, 2, 5, 3, and 2 trials of Blocks 1 through 7, respectively. The apparent sharp improvement in the performance of the two unstructured groups in the final acquisition block (5) was probably due to strategy overlap, since the color and shape strategies coincided on an unusually large proportion (.50) of the trials in the last acquisition series.

To eliminate shape-color overlap as a source of confounding, a corrected score was obtained by (a) removing from the analysis those trials

on which strategies coincided, then (b) calculating, for each subject, the proportion of "color" strategy responses in the remaining trials of the block. The means of the resultant scores are shown in Table 1. An analysis of the corrected acquisition scores yielded all of the significant effects found in the previous analysis (information, trials, and Trials \times Information, p 's $< .001$). In addition, two interactions involving stimulus type were significant, that is, Stimulus Type \times Information, $F(1, 36) = 4.76$, $p < .05$; Stimulus Type \times Information \times Trials, $F(4, 144) = 3.55$, $p < .01$. Reference to Table 1 suggests that real differences in the effectiveness of the verbal and nonverbal stimulus events were obscured in the initial analysis. Subjects presented with an unstructured verbal stimulus showed some improvement over the five blocks of the acquisition phase, while those presented with the unstructured nonverbal stimulus demonstrated scant evidence of learning. Further analysis of the Stimulus Type \times Information interaction by the Newman-Keuls procedure indicated that, with a single exception, all differences between pairs of acquisition means of the four experimental conditions were statistically significant ($p < .01$). Only the difference between the two structured groups was not reliable.

In summary, the analysis of the corrected

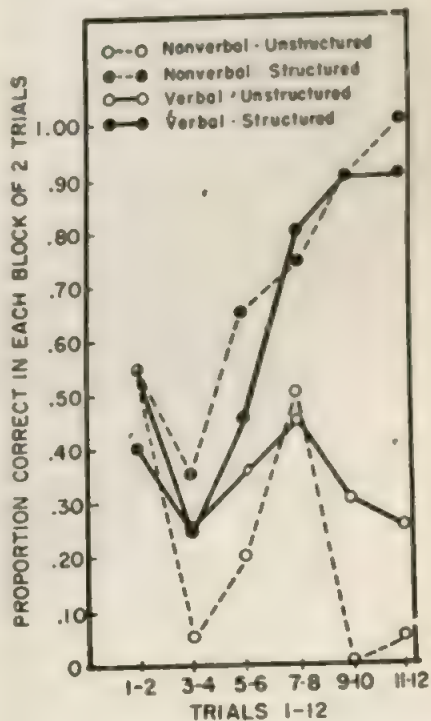


FIG. 2. The mean performance of the four experimental groups during the initial phase of acquisition (Trials 1-12).

TABLE 1

MEAN PROPORTION CORRECT IN THE FIVE ACQUISITION AND THE TWO EXTINCTION BLOCKS AFTER REMOVAL OF "OVERLAPPING" TRIALS

Condition	Blocks of 10 trials						
	Acquisition					Extinction	
	1	2	3	4	5	6	7
Structured Verbal	.56	.91	.93	.96	.92	.71	.68
Nonverbal	.57	.91	1.00	1.00	1.00	.76	.66
Unstructured Verbal	.28	.32	.40	.48	.56	.50	.46
Nonverbal	.08	.10	.11	.14	.14	.09	.11

acquisition scores indicated that subjects presented with the unstructured verbal event performed at a significantly higher level than subjects presented with the unstructured nonverbal event. Both groups of subjects, however, were strikingly inferior in acquisition performance to subjects given the brief informational set.

Extinction Performance

To determine the effects of extinction, the performance of subjects during the two nonreinforcement blocks was compared with their performance in the last acquisition block. Because the number of correct responses in the final acquisition block was spuriously inflated for all subjects who had followed a shape strategy, the analysis was performed on the corrected scores summarized in Table 1. Analysis of these data for changes as a function of trials indicated a significant trial-block effect, $F(2, 72) = 6.49$, $p < .01$, and an Information \times Trials interaction, $F(2, 72) = 6.48$, $p < .01$. Neither the Stimulus Type \times Trials nor the three-factor interaction was statistically reliable ($p > .10$). It should be noted, however, that one of the unstructured groups (nonverbal) obtained a very low mean score in Block 5. A large decrement in performance during extinction for this group was impossible, and, therefore, the interpretation of the Trials \times Information interaction is unclear.

Inspection of Table 1 suggests that marked differential effects appeared early in the extinction phase. Separate within-group analyses of the performance shifts that occurred between the last acquisition block and the first extinction block support this observation. Subjects in the structured verbal condition produced significantly fewer color strategy responses in Block 6 than in Block 5, $t(9) = 3.11$, $p < .02$. A similar decrement was obtained in the structured nonverbal condition, $t(9) = 3.29$, $p < .01$. However, in the case of both unstructured conditions, Block 5 performance could not be reliably differentiated from that of Block 6 or Block 7.

Additional information on the effects of nonreinforcement for subjects in the unstructured nonverbal condition is given by an inspection of their individual protocols. Within this group, every subject continued to follow throughout extinction the response strategy (e.g., color, shape) that he had followed in the final acquisition block.

Sex was not significant, either as a main effect or in interaction with the treatment factors, in the acquisition or nonreinforcement results.

DISCUSSION

The proposal that the reinforcement properties of any event vary as a function of its informational characteristics is not a new one (Thordike, 1938). It has, however, been given little attention in recent discussions of the motivational determinants of the effectiveness of verbal stimuli. The present data, while they do not necessarily contradict the motivational proposals, underline the importance of considering the cue properties of the reinforcement stimulus. Furthermore, it seems reasonable to suggest, as a hypothesis for further study, that the various motivational operations that have been reported in the recent literature are effective because they indirectly structure the signal properties of the stimulus.

One of the more interesting outcomes of this study was the extent to which the word "right" was influenced by the instructional set. Recent evidence indicates, contrary to what has been commonly assumed, that "right" (undefined) is a relatively ineffective reinforcer for children in the 6-12-year-old age range (Curry, 1960; Larsen, 1963). It is not that children, when asked, are unable to provide an appropriate definition of the word. Rather, it appears that they fail to make the linkage between this expression of the experimenter and the adequacy of their choices on the criterion task. Moreover, even among adult subjects, common expressions of general agreement do not invariably constitute a reinforcing state of affairs (e.g., Hildrum & Brown, 1956; Mandler & Kaplan, 1956). These findings are consistent with the suggestion that words of assent approval, perhaps as a result of their commonplace and nondiscriminant usage, are relatively ambiguous events for many subjects when they are introduced in the laboratory to signal the correctness or incorrectness of a specific action.

Finally, it should be noted that the extinction data indicated that the nonoccurrence of an informative event was associated with a shift in response strategy. This finding appears to contradict the "blank trials law" recently proposed by Levine, Leitenberg, and Richter (1964). According to this theorem, when the experimenter says nothing at all (i.e., a "blank" trial), sub-

jects should behave as though the experimenter were saying "right." However, in the present experiment, the omission of "right" (or the non-verbal event) in the structured groups was associated with an exploration of alternative strategies; obviously the experimenter's silence was not equivalent to the experimenter's saying "right." While the present study differed from those of Levine et al. (1964) in a number of respects (age of subjects, tasks employed, etc.), one of the possibly crucial differences was that subjects in this experiment were not forewarned that the experimenter-produced outcomes would be discontinued. In the experiments of Levine et al. (1964), subjects were forewarned. If further work demonstrates that such additional information is essential to the blank trials law, its generality would be markedly restricted. On the other hand, it seems reasonable to propose that the experimenter's silence, like other nonverbal events, can function as a positive, negative, or neutral signal, according to its definition within the experimental context.

REFERENCES

- CURRY, C. The effects of verbal reinforcement combinations on learning in children. *Journal of Experimental Psychology*, 1960, 59, 434.
- EGGER, M. D., & MILLER, N. E. Secondary reinforcement in rats as a function of information value and reliability of the stimulus. *Journal of Experimental Psychology*, 1962, 64, 97-104.
- GEWIRTZ, J. L., & BAER, D. M. Deprivation and satiation of social reinforcers as drive conditions. *Journal of Abnormal and Social Psychology*, 1958, 57, 165-172.
- GRANT, D. A., & BERG, E. A. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology*, 1948, 38, 404-411.
- HILDRETH, D. C., & BROWN, R. W. Verbal reinforcement and interview bias. *Journal of Abnormal and Social Psychology*, 1956, 53, 108-111.
- LARSEN, C. C. A comparison of the effects of verbal and non-verbal reinforcements on learning in children. Unpublished doctoral dissertation, Indiana University, 1963.
- LEVINE, M., LEITENBERG, H., & RICHTER, M. The blank trials law: The equivalence of positive reinforcement and nonreinforcement. *Psychological Review*, 1964, 71, 94-103.
- MANDLER, G., & KAPLAN, W. K. Subjective evaluation and reinforcing effect of a verbal stimulus. *Science*, 1956, 124, 582-583.
- PARTON, D. A., & ROSS, A. O. Social reinforcement of children's motor behavior: A review. *Psychological Bulletin*, 1965, 64, 65-73.
- POSTMAN, L., & SASSENATH, J. The automatic action of verbal rewards and punishments. *Journal of General Psychology*, 1961, 65, 109-136.
- SPENCE, J. T. Verbal-discrimination performance as a function of instructions and verbal-reinforcement combination in normal and retarded children. *Child Development*, 1966, 37, 269-281.
- STEVENSON, H. W. Social reinforcement with children. In L. Lipsitt & C. C. Spiker (Eds.), *Advances in child development and behavior*. Vol. 2. New York: Academic Press, 1965. Pp. 97-126.
- THORNDIKE, E. L. The law of effect: A round table discussion. III. *Psychological Review*, 1938, 45, 204-205.
- WALTERS, R. H., & PARKE, R. D. Social motivation, dependency, and susceptibility to social influence. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. New York: Academic Press, 1964. Pp. 231-276.

(Received February 14, 1966)

Journal of Personality and Social Psychology
1967, Vol. 5, No. 3, 357-363

EFFECTS OF THE PARTNER'S ABRUPT STRATEGY CHANGE UPON SUBJECT'S RESPONDING IN THE PRISONER'S DILEMMA¹

PAUL G. SWINGLE

AND

HENRY COADY

McGill University

Dalhousie University

Male Ss initially exposed to either a 100% cooperative or 0% cooperative partner for 50 trials in a Prisoner's Dilemma game were exposed, during trials 51-100, to a partner whose strategy was either 0%, 25%, 50%, or 100% cooperative. Strategy change resulted in increased variability in S's strategy although mean number of cooperative responses was not significantly affected. An uncooperative strategy, at either stage of the game, resulted in a less favorable evaluation of the partner.

Although maximum joint payoff in the Prisoner's Dilemma (PD) game is achieved by mu-

tual cooperation, investigators have repeatedly found a relatively low level of cooperation. Sev-

¹This research was supported by National Research Council of Canada Grant APA-158 to the first author. The authors are indebted to the staff

at HMCS Cornwallis for their cooperation and support during the course of this study. Particular thanks are extended to C. Bailey and F. Came.

eral antecedent conditions which have been found to enhance cooperative responding include cooperative orientation (Deutsch 1958), enhancing the payoff for mutual cooperation (Minas, Scodel, Marlowe, & Rawson, 1960) or unilateral cooperation (Sermat, 1964), reversible choice (Deutsch 1958), public commitment to cooperation (Oskamp & Perlman, 1965), and prior social interaction between subjects (Scodel, 1963).

A factor of particular interest in the study of competition is the strategy of the partner. Solomon (1960) in a sequential move PD game found about 29% cooperative responding by subjects when the partner was unconditionally cooperative, and Minas et al. (1960) found about 40% cooperation when the partner cooperated 100%.

The effects of various strategy mixtures have also been compared. McClintock, Harrison, Strand, and Gallo (1963) compared subjects playing against programmed partners whose level of cooperative responding was either 85%, 50%, or 15% and found no significant differences. Bixenstine and Wilson (1963) found little initial difference when the opponent was programmed for 5% or 95% cooperation.

Sequential strategies have been found to influence the subject's level of cooperative responding. Bixenstine and Wilson (1963) found that sequences from 5% to 95% cooperation did result in a significant increase in the subject's cooperativeness, and Scodel (1962) found that 0-100% shifts in strategy by the programmed partner result in increased cooperative responding by the subject.

The study of the effects of strategy change upon cooperative responding involves, first, a consideration of the nature of the strategy shift and, second, a consideration of the conditions preceding the change.

The abruptness or distinctiveness of any change in strategy depends upon the overall size of the change and length of time over which the change takes place. The most abrupt strategy change, from one pure strategy to the alternative pure strategy (e.g., 100-0% cooperation), may be made less distinctive if the change occurs gradually over a long series of trials. A second factor which should be important is the prechange strategy of the opponent and the length of time the antecedent strategy was in effect.

The present experiment compared the levels of competitive responding and the "time to decision" of four different levels of opponent strategy which were abruptly "adopted" after 50 trials of either an unconditionally cooperative

or unconditionally uncooperative strategy.

METHOD

Subjects

Eighty male recruits taking basic training at HMCS Cornwallis naval training base were subjects.

Overview of the Experimental Situation

Subjects were seated in well-ventilated 4 X 6 cubicles, acoustically isolated from both the other subjects and the programming and recording apparatus. Four such cubicles were located in a room, with the result that subjects had no idea exactly whom they might be playing against. The cubicle contained an intercom system and a console. The console consisted of two response levers (momentary toggle switches), a "go" light, a display window, and a counter which automatically recorded the subject's cumulative score. The PD matrix shown in Figure 1 was affixed to the display window.

Although firmly convinced that one of the other persons in the experimental room was their opponent, subjects were actually playing against a programmed apparatus. A probability generator was used to simulate the opponent's strategy.

Procedure

Subjects usually arrived in groups of four and typically waited in a room provided for this purpose for several minutes until earlier subjects were finished. No attempts were made to discourage interaction. Subjects were then taken one by one and seated in the isolation cubicles. After all four were seated, tape-recorded instructions were presented via the intercom system. As two experiments were being run simultaneously, subjects at times had to wait in the cubicles several minutes before they received the instructions.

The instructions and reward structure were designed to encourage competitive behavior. Subjects were told that the experimenters were "concerned with how people make simple decisions" and that "to make it interesting, we have made it into a game." After a detailed description of the game

		PLAYER B (partner) y	
		x	
PLAYER A (subject)	1	<div style="display: inline-block; text-align: center;">3 / 3</div>	<div style="display: inline-block; text-align: center;">4 / 0</div>
	2	<div style="display: inline-block; text-align: center;">4 / 0</div>	<div style="display: inline-block; text-align: center;">1 / 1</div>

FIG. 1. Payoff matrix for the PD game. (The numbers in the shaded half of each quadrant represent the subject's payoff.)

subjects were told that "the one person who, at the end of the game, has the most points shown on his counter will have a chance to win several dollars. Every day we award several dollars to our highest scorers."

After questions were answered, the game started. The red "go" light went on, indicating the start of a trial to the subject. The subject made his response by pushing either lever Number 1 or 2, and after a delay of 3 seconds a light came on behind the appropriate quadrant of the matrix affixed to the display window. During this 6-second visual display, the subject's points were automatically cumulated on his counter. The intertrial interval, therefore, was about 9 seconds.

Time to decision was obtained by having the "go" light onset activate a timing device which was stopped by the subject's response.

Eight groups of subjects were combined in a 2×4 factorial design. Four groups of subjects received 50 trials with the simulated partner set at 0% cooperation (Condition A).² At Trial 51, each group experienced a different treatment condition: A₁B₁ experienced the most radical change, a change to 100% cooperation; A₁B₂ was changed to 50% cooperation; A₁B₃ was changed to 25% cooperation; and A₁B₄ remained unchanged at 0% cooperation.

The four remaining groups received 50 trials with a simulated 100% cooperative partner (Condition A₂). At Trial 51, Group A₂B₁ remained unchanged at 100%; A₂B₂ was changed to 50%; A₂B₃ to 25%; and A₂B₄ to 0%.

In total, the subjects were run for 100 trials, 50 under Condition A and 50 under Condition B.

At the end of the experiment, subjects were asked to complete two questionnaires. The first simply asked whether the subject would want the same partner if he were to serve again as a subject. A 4-point scale was provided: "yes," "maybe," "rather not," "no." The second scale required the subject to give his "best guess" about the person against whom he was playing. The scale consisted of 12 evaluative adjectives, 6 positive (pleasant, kind, etc.), and 6 negative (selfish, rude, etc.), all of which were to be answered by a "yes" or "no" reply.

The recording of decision time and noncooperative responses, as well as trial pacing, strategy changes, feedback, etc., were all accomplished by means of standard relay circuit programming and recording equipment.

RESULTS

The competitive response data for Conditions A₁ and A₂ are shown in Figures 2 and 3. The experimental groups are combined in each figure.

² Checks for \$2 were sent to about seven subjects per week.

³ The levels of cooperativeness, 0%, 25%, 50%, and 100% are nominal values. The actual levels for the 0% and 100% were about 4% and 96%, respectively.

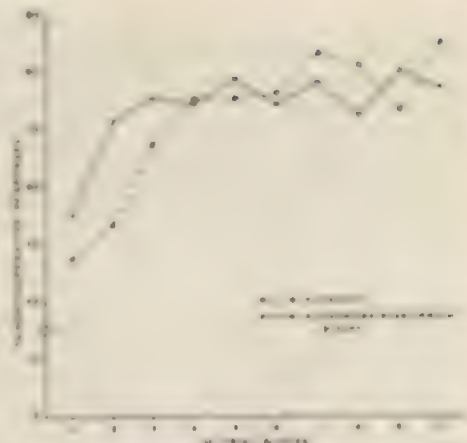


FIG. 2. Comparison of the subject's mean percentage of noncooperative responses between the three combined experimental groups exposed to an initially cooperative partner and the control group exposed to a 100% cooperative partner over the entire 100 trials of the game.

In Condition A₂, trials of unconditional noncooperation (preventing partner's strategy change) there is very little difference between the combined experimental groups' curve and that of the control group (100 trials of unconditional noncooperation from partner). At Trial Block 6, the experimental groups show a slight tendency to reduce the number of competitive responses.

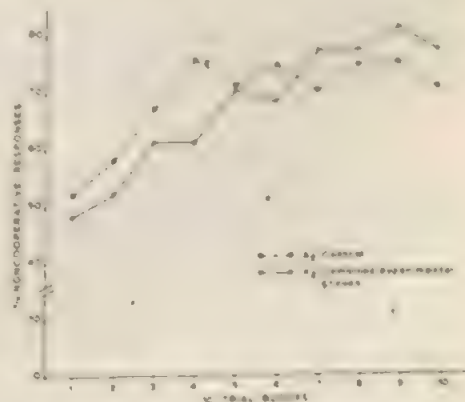


FIG. 3. Comparison of the subject's mean percentage of noncooperative responses between the three combined experimental groups exposed to an initially cooperative partner and the control group exposed to a 100% cooperative partner over the entire 100 trials of the game.

In Condition A_2 (50 trials of 100% cooperation) the difference between experimental and control groups is also slight and, as in A_1 , the experimental groups show a slight reduction in the number of competitive responses at Trial Block 6. The A_2 experimental groups differ from the A_1 experimental groups, as indicated by the curves in Figures 2 and 3, in that the number of competitive responses for combined A_1 experimental groups rises rapidly and asymptotes at about 75% by Trial Block 3, whereas the A_2 curve indicates a slower but continuous rise in competitive responding, reaching a similar level only after 70 trials.

The response data were transformed to arc sines, and 2×4 analyses of variance were computed on the data for the first five trials, the last five trials, Trial Block 5, Trial Block 6, and the differences between Trial Blocks 5 and 6 (i.e., number of competitive responses at Trial Block 6 less the number at Trial Block 5). No significant main or interaction effects were obtained. The interaction between the A and B effects did approach significance in the last five-trial data ($P = 2.00$, $.15 > p > .10$). Inspection of the data, as shown in Figure 2, indicated that there was a tendency for positive strategy changes (i.e., partner is initially 0% cooperative and at a later stage increases his level of cooperative responding) to result in greater cooperation than in the control condition (i.e., 100 trials at 0% cooperation), while the reverse seems to be the case when the partner is initially 100% cooperative (Figure 3).

The lack of significance of the difference data, even though large changes in responding were observed, indicated that the direction of change (i.e., more or less cooperative responses after the change in the partner's strategy) was not con-

TABLE 1

ANALYSIS OF VARIANCE OF ABSOLUTE VALUE OF DIFFERENCE IN SUBJECT'S COMPETITIVE RESPONDING FOLLOWING PARTNER'S STRATEGY CHANGE AS A FUNCTION OF PARTNER'S INITIAL AND FINAL STRATEGY LEVELS

Source	SS	MS	F
A (partner's initial strategy)	.04	.04	
B (partner's final strategy)	1.56	.52	5.20**
A \times B	.84	.28	2.80*
Error	7.26	.10	
Total	9.70		

* $p < .05$, $df = 3/72$.

** $p < .01$, $df = 3/72$.

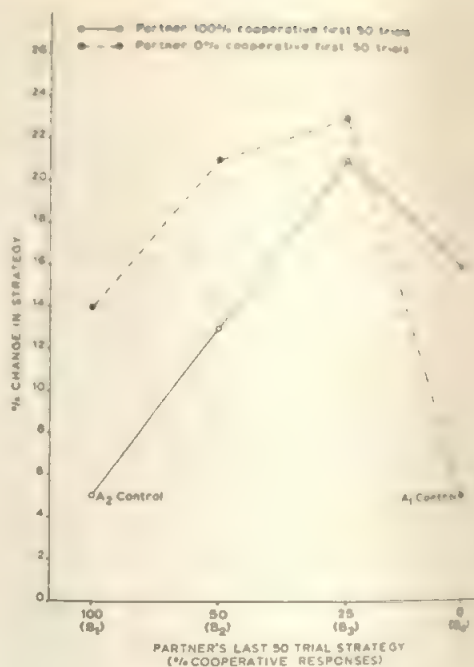


FIG. 4. Average percentage of change in the subject's strategy resulting from abrupt strategy changes by the partner as a function of the prechange and postchange strategy of the partner.

sistent and that the large changes, being in both directions, were canceling one another out, with the result that a small mean difference but a large error term was obtained. To test the notion that change in partner's strategy does result in large individual shifts in the subject's strategy, the direction of which is, in part, determined by the prechange level of cooperation, an analysis of variance (see Table 1) was computed on the arc sine of the absolute value of the difference data (i.e., the difference between the number of noncooperative responses at Trial Block 6 and the number at Trial Block 5, sign ignored).

The analysis indicates a significant main effect due to the last 50-trial strategy of the partner and a significant interaction between pre- and postchange strategies. The percentage of change in each condition is shown in Figure 4.

The amount of variation in the control conditions is identical and negligible, 5% in each condition. All experimental conditions appear to be markedly more variable than the controls, the greatest absolute change occurring when the partner shifts to a 25% level of cooperation.⁴

⁴ It should be noted that although the cooperative levels in the 25% condition are identical during

The differences between the 100% shift experimental groups (A_1B_1 , A_2B_4) and either control group (100 trials at 0% or 100% cooperation) are significant by t test beyond the .05 level. More conservative comparisons of each experimental group versus its respective control by the Dunnett test (Winer, 1962) indicated that A_1B_1 and A_1B_2 differed from A_1B_4 ($p < .05$, $p < .05$, respectively) and that the difference between A_1B_1 and control approached the conventional level of significance ($.10 > p > .05$). In Condition A_2 , the differences between experimental groups and control were significant in Groups A_2B_3 ($p < .005$) and A_2B_4 ($p < .05$) and approached significance in Group A_2B_2 ($.10 > p > .05$).

The data as shown in Figure 4 suggest that greater change in the subject's responding resulted when the partner's strategy shifted from 0% cooperation (i.e., positive change) than when the shift was from an initially 100% cooperative strategy (i.e., negative change). The $A \times B$ interaction in the analysis of variance (Table 1) indicates that the observed difference may be

trials 51-100, the amount of strategy shift is not the same. That is, the shift from 100% to 25% is nominally considerably greater than the shift from 0% to 25%.

significant. A t test between A_1 experimental groups combined (Group A_1B_1 , dropped) and A_1 experimental groups combined (A_1B_2 , dropped) indicated that the difference is reliable (t value, $p < .01$).

The decision-time latency data are shown in Figure 5. Four curves are shown, one for all experimental groups in Condition A_1 combined, one for all experimental A_2 groups combined, and one for each control group.

The curves show the expected rapid decline to a steady level after a few trials. Analysis of variance of the temporal of the latency series indicated no difference due to the B effect³ for the first five trials, the last five trials, Trial Block 5, Trial Block 6, nor the difference in

Although the curves suggest an initial difference between experimental and control groups prior to the differential treatment commencing at Trial 51, the differences result from a few subjects with exceedingly large latencies. Mann-Whitney U tests between each experimental group, separately, and the appropriate control group for the first 50 trial latency totals indicated no significant differences ($\alpha = .05$). There were no procedural irregularities nor other nonrandom factors, to the authors' knowledge, which would differentially affect the groups prior to the introduction of strategy shifts at Trial 51.

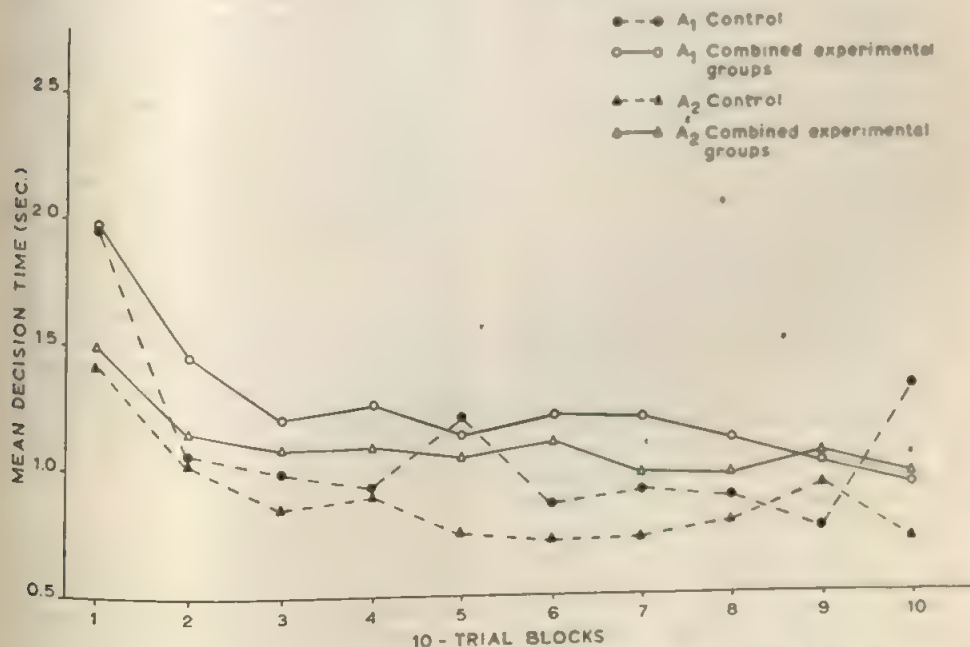


FIG. 5. Comparison of the time courses of the decision time between the experimental and control groups distinguished on the basis of the partner's initial strategy level.

TABLE 2

ANALYSIS OF VARIANCE OF SUBJECT'S EVALUATIVE RATINGS OF PARTNER AS A FUNCTION OF PARTNER'S INITIAL AND FINAL STRATEGY LEVELS

Source	SS	MS	F
A (partner's initial strategy)	3.65	1.22	8.13**
B (partner's final strategy)	.64	.64	4.27*
A × B	.36	.12	
Error	10.87	.15	
Total	15.52		

* $p < .05$, $df = 1/72$.

** $p < .01$, $df = 3/72$.

speed after the strategy change (Trial Block 6 minus Trial Block 5 speed scores). There was a significant A effect for the first five trials, indicating that subjects respond more slowly when confronted with a partner whose beginning game strategy is 0% cooperative ($F = 4.75$, $p < .05$). This difference was not significant at Trial Blocks 5 or 6, or during the last five trials, and the difference scores (Trial Block 6 minus Block 5 speed scores) were also nonsignificant.

The major problem with the decision-time data was the extreme variability within subjects. The basic assumption was that the subject's decision time would increase following an abrupt strategy change by the partner. This assumption is supported, since 39 of the 60 experimental subjects' decision times increased in Trial Block 6 relative to Trial Block 5 ($p = .014$, one-tailed), whereas only 9 of the 20 control subjects' decision times increased ($p = .412$).

Attitude Measures

The scale designed to assess the subject's willingness to confront his present partner in any future games was scored on a 1 through 4 scale: "yes" (would be willing to play against partner in the future) scored as 4; "maybe," 3; "rather not," 2; "no," 1. Analysis of variance indicated that the beginning strategy level treatment effect (A effect) was significant ($p < .05$). The average rating was 3.22 for subjects playing against initially uncooperative partners, and 3.55 for subjects having initially 100% cooperative partners. Neither the cooperative level following strategy change (B effect) nor the interaction was significant.

The scale designed to assess the subject's evaluation of the partner was scored on a 0 through 12 basis, where 12 represented the most favor-

able attitude toward the partner. The analysis of variance of the arc-sine transformed data is shown in Table 2. The significant A effect is in agreement with the replay scale in that initially cooperative partners are evaluated more favorably ($\bar{X} = 10.94$) than are the initially uncooperative partners ($\bar{X} = 10.10$).

The significant B effect indicates that the level of the partner's cooperation following the 50-trial condition affects the resulting evaluation independently of the A effect (i.e., no A × B interaction). Application of the Newman-Keuls studentized range statistic (Winer, 1962) indicated that the 0% level of cooperation was significantly different from all other conditions ($p < .01$), but the remaining three levels were not significantly different. Thus, there is a consistency in both scales in that the 0% cooperative level seems to be directly related to negative evaluation.

DISCUSSION

The present study was designed to determine the effects of the partner's early strategy and subsequent abrupt change in strategy in a PD game situation upon the subject's cooperative responding, decision time, and attitude toward the partner. The experimental literature on the effects of the opponent's strategy upon the subject's cooperative behavior indicates that the subject tends to be rather insensitive to the opponent's level of cooperation. Abrupt changes in strategy such as 0-100% cooperation (Schoedel, 1963) and 5-50-95% cooperation and the reverse sequence (Bixenstine & Wilson, 1963) have been found to give rise to changes in the subject's cooperative behavior, although no effect of strategy change upon mean levels of cooperative responding was found in the present study.

The present finding of no significant difference in the number of cooperative responses between groups exposed to an unconditionally cooperative partner versus those exposed to an unconditionally uncooperative partner during the first 50 trials is in agreement with previous research (Bixenstine & Wilson, 1963; McClintock et al., 1963).

The major effect of the strategy change was that of increased variability in the subject's responding following the change. This indicates that comparing the average number of competitive responses between groups exposed to different strategy manipulations may not be a sensitive index of the treatment effect. This seems

reasonable if one considers the possibility of response limits in non-zero-sum game situations. If, for example, a subject was responding at or near a pure strategy level (i.e., about 100% cooperative or uncooperative responses), then any demonstrable effect of a strategy change would be limited to increased choice of the previously unused response. Further, a subject responding at or about the 50% level (i.e., 50% cooperative and 50% uncooperative responses) might be assumed to be attempting to mildly exploit an unconditionally cooperative partner or, conversely, to be attempting to persuade an unconditionally uncooperative partner to become more cooperative by himself making some cooperative gestures. If this is the case, then the experience of an abrupt change in the partner's strategy might, with more or less equal frequency, result in a negative change (e.g., the subject, angered at the partner's strategy change, makes more Type 2 responses) or a positive change (e.g., the subject believes the partner's strategy change is a just retaliatory move precipitated by the subject's exploitative responses, so that the subject increases his cooperative responses to encourage the partner to again cooperate). The results of the present experiment support the above argument that response variability, or simply magnitude of strategy change, is a more sensitive measure of treatment effects.

The decision-time data were rather disappointing. Although the decision time did show a slight tendency to increase following any strategy change, the measure is not sensitive to the extent or direction of the partner's strategy change. The fact that decision time was significantly greater in the initial trials for subjects exposed to an unconditionally uncooperative partner suggests that subjects in this situation find it more difficult to decide how to deal with their partner.

The postexperimental rating scales indicate that the unconditionally uncooperative partner is less favorably evaluated and tends to be less desirable as a future partner. This effect maintains regardless of when the partner was uncooperative—partners who are 100% uncoopera-

tive during either of the two phases. Blocks are negatively evaluated.

The results of this experiment indicate that further study of the factors which influence the stability of the partner's strategy change would be expected. It would also be of interest to investigate the factors which give rise to differences in individual subjects' responses to partner's strategy changes.

REFERENCES

- BIXENSTINE, V. F., & WILSON, K. V. Effects of level of cooperative choice by the other player on choices in a Prisoner's Dilemma game: II. *Journal of Abnormal and Social Psychology*, 1963, 67, 139-147.
- DITTMER, M. Trust and response. *Journal of Conflict Resolution*, 1966, 4, 161-172.
- MCCLINTOCK, C. G., HARRISON, A., STANLEY, S., & GATTO, P. Information and decision strategy of the other player, and two-person game behavior. *Journal of Abnormal and Social Psychology*, 1963, 67, 631-635.
- MINAS, J. S., SCODEL, A., MARLOWE, D., & RAWSON, H. Some descriptive aspects of two-person non-zero-sum games: II. *Journal of Conflict Resolution*, 1960, 4, 193-197.
- OSKAMP, S., & PERLMAN, D. Factors affecting cooperation in a Prisoner's Dilemma game. *Journal of Conflict Resolution*, 1965, 9, 359-374.
- SCODEL, A. Induced collaboration in some non-zero-sum games. *Journal of Conflict Resolution*, 1962, 6, 335-340.
- SCODEL, A. Effects of group discussion on cooperation in a Prisoner's Dilemma game. Paper read at California Psychological Association, San Francisco, December 1963.
- SERMAT, V. Cooperative behavior in a mixed-motive game. *Journal of Social Psychology*, 1964, 62, 217-239.
- SOLOMON, L. The influence of some types of power relationships and game strategies upon the development of interpersonal trust. *Journal of Abnormal and Social Psychology*, 1960, 61, 223-230.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received April 12, 1966)

STIMULUS QUALITIES OF THE TARGET OF AGGRESSION: A FURTHER STUDY¹

LEONARD BERKOWITZ² AND RUSSELL G. GEEN

University of Wisconsin

Each of 90 male Ss was first deliberately provoked by E's accomplice, then watched either (a) justified film aggression, (b) less justified film aggression, or (c) an exciting but nonaggressive track race. He was then given an opportunity to administer electric shocks to the accomplice, after learning the accomplice's name was either "Kirk" or "Bob." For the men seeing the aggressive film, the name "Kirk" associated the accomplice with the victim of the witnessed violence. Even though the available target's name was introduced after the film was seen, the target was attacked more often when inhibitions against aggression were weakest and the target's name associated him with the victim of the movie aggression. As in other experiments, the target's cue value for aggression determined the magnitude of aggression directed against him.

What determines the intensity of the aggression that is directed against a particular target? Most answers to such a question undoubtedly would attribute the strength of a given attack primarily to internal factors, such as drive or emotion arousal, previously acquired aggressiveness habits, and/or inhibitions against this form of behavior. External determinants must also be considered, however. Continuing an argument advanced in a number of papers (Berkowitz, 1962, 1964, 1965b), we will here attempt to show that the stimulus characteristics of the available target can also govern the magnitude of the attack.

In two recent experiments (Berkowitz & Geen, 1966; Geen & Berkowitz, 1966) male college students were first made angry with a peer and then were shown a brief film. Some of these people watched a fairly violent prize-fight scene, while other subjects saw an exciting but nonaggressive track race. Then, immediately after the film, each subject was provided with a socially sanctioned opportunity to give electric shocks to the person who had provoked him. The main hypothesis underlying this research contends that a target will elicit aggressive responses from an individual who is ready to act aggressively to the extent that the target is associated with aggressive behavior generally. This association was established in some conditions by giving the anger instigator the same name as the victim of the filmed violence, either the name of the actor portraying the victim (Berkowitz & Geen, 1966) or the fictitious name used in the story (Geen & Berkowitz,

1966). In both instances the greatest number of shocks was given to the anger instigator whose name connected him with the victim of the witnessed aggression, while such naming had no effect on the people seeing the nonaggressive movie.

Although highlighting the importance of the available target's stimulus characteristics, the findings in these two studies are open to a number of alternative explanations. In this paper we will focus our attention on a possible explanation suggested by the Spence (1956) incentive conception. Adherents of this mode of thought would (or should) point to the experimental procedure followed in establishing the target person's name-mediated association with the film victim; in both experiments the confederate was introduced *before* the movie was shown. Seeing a film character with the same name as the anger instigator conceivably could have reminded the angered subjects of the provocation they had suffered, so that, in essence, they were kept aroused by the film, while the subjects in the other conditions were permitted to calm down. Other evidence indicates, however, that our findings are not due to a simple reminder of the anger instigator. In an unpublished experiment by the present writers, deliberately angered subjects were shown the exciting but nonaggressive track-race film employed in this research program after having been led to believe their tormentor had the same name as either the winner or the loser in the race. Even though the subjects, hearing this name, could have been reminded of the anger instigator, the men in these conditions did not give their frustrater more shocks immediately afterwards than did the people in the control groups. The name evidently had to appear in an aggressive context if it was to have an effect upon subsequent aggression.

¹This study was carried out by RGG under LB's supervision as part of a project sponsored by Grant G-23988 from the National Science Foundation to the senior author.

²Now at Cornell University.

Spence's incentive formulation can readily account for these results, however. When the subject saw a person beaten who had the same name as his own frustrater, he could have been stimulated to make anticipatory aggressive responses. The excitation produced by this anticipatory aggression then might have led to the great number of attacks upon the frustrater soon afterwards.

The present experiment was designed as a test of this alternative, "anticipation" explanation of the earlier results. Subjects were angered and then shown the prize-fight film, as in the Berkowitz-Geen and Geen-Berkowitz studies, but this time they were not told the frustrater's name until the movie had ended. If, in the earlier studies, the film victim's name had elicited anticipatory aggression as the subjects watched the prize-fight scene, this altered procedure should now eliminate or lessen the effects of the available target's name. On the other hand, greater aggression should be directed against the frustrater having the same name as the victim of the witnessed violence if this label increases the frustrater's cue value for aggression.

Subjects

The subjects were 90 male undergraduates at the University of Wisconsin who volunteered without knowing the purpose of the experiment in order to earn points counting toward their final grade in introductory psychology.

Method

When each subject, scheduled singly, came to the laboratory, he was met by the experimenter (RGG) and a confederate who was posing as another subject. The experimenter told the two men the study had to do with problem solving under stress. He said the stress would be induced by administering mild electric shocks and gave the subject an opportunity to withdraw from the experiment if he so desired. (None of the people did so, although one person left later after revealing he had heard about the investigation.) The two men then went to separate rooms where they were to hear further instructions over an intercom. Actually delivered by a tape recording, the oral instructions given to the subject informed him that he would have to work on a problem while his partner in the other room watched a brief film. In about 5 minutes, at the conclusion of the film, the partner would judge the quality of the subject's problem solution. These evaluations were to be administered in the form of electric shocks; one shock for an excellent solution, up to 10 shocks for a very poor solution.

The subject then was given the specific problem he was to solve: to devise a contest that would be part of an advertising campaign for a store. He was also told at this time that his partner's later task would be to formulate a promotion campaign for a new laundry detergent. At the end of the tape recording, the experimenter came into the subject's room and gave him a pad of paper onto which he was to write his solution to the assigned problem.

Five minutes later the experimenter returned, brought a film camera into the subject's room, and placed up the subject's work, supposedly to take it to the other person for evaluation. In actuality, however, the experimenter went to a nearby room from which he gave the subject seven shocks in order to produce a moderate emotion arousal in him. He then went back into the subject's room and asked him to complete a brief questionnaire assessing his present mood.

When this was done the experimenter introduced the film the subject was to see as part of the next phase of the study. Three film conditions were established. Two-thirds of the cases (*aggressive film groups*) were to be shown the prize-fight scene used in all of the experiments in this research program, but the introduction to that film was varied. The experimenter played a tape recording verbally summarizing the film story up to the witnessed scene, ostensibly so that the subject would have a better understanding of the scene. For one *aggressive film* condition this story synopsis portrayed the movie protagonist in a very unfavorable light so that the subject would regard the beating the protagonist was to receive in the film as relatively proper (*justified-film-aggression* condition). The other *aggressive film* groups were given a story depicting the protagonist in a more favorable manner. Being more sympathetic to the protagonist, the men in this condition presumably would regard the beating the hero took in the movie as *less-justified film aggression*. The remaining third of the cases saw the exciting but nonaggressive track-race film used in the earlier experiments (*track-race-film* condition). Previous investigations in this series (Berkowitz, 1965a, Berkowitz, Corwin, & Heronimus, 1965; Berkowitz & Rawlings, 1955) have demonstrated that the portrayal of the film protagonist as a scoundrel by means of the *justified aggression* introduction effectively lowered the subject's inhibitions against aggression on his part immediately after the film. The *less-justified film aggression* introduction was added in order to determine whether the accomplice's name would govern the magnitude of the attacks made upon him even under relatively strong restraints against aggression.

At the conclusion of the film the subject filled out a second mood questionnaire. The confederate, who supposedly had been working on the problem assigned to him, then entered the room saying he had finished his task. The experimenter took this person's problem solution and then set the stage for the final experimental manipulation. Saying he wanted to make sure he did not get their records mixed up, he asked the two men for their names. After the subject had given his name, within each of the film conditions the confederate said his name was either *Kirk Anderson* or *Bob Anderson*. (If the subject had witnessed the prize-fight scene in which the actor, Kirk Douglas, played the role of the person taking the beating, and the confederate had just introduced himself as "Kirk," the experimenter made a remark about the coincidence in the names. When queried at the end of the session, 4 subjects indicated they had become suspicious at this point, and they are not included in the 90 subjects. None of the other men said they had doubted the coincidence

after the deceptions were explained at the conclusion of the session.)

The two men were sent to separate rooms again, and about 30 seconds later the experimenter gave the subject what was supposedly his partner's solution to his problem. The subject was reminded that he was to give his partner from 1 to 10 shocks as his evaluation of his partner's solution, and also that this was the last time shocks were to be administered. After providing this information, the experimenter left the room so that the subject would be alone while giving the shocks. Although these shocks serve as the principal measure of aggression, a final questionnaire was also employed for secondary measures. The subject completed this questionnaire shortly after giving his "evaluation" of the other person's problem solution, indicating his attitude toward his partner on four 7-step scales. At the conclusion of the session the deceptions were explained, and each subject was asked not to talk about the experiment for the remainder of the semester.

RESULTS

Film-Induced Mood Changes

The first question to be considered has to do with the effects of the film treatments on the subjects' experimental mood. As was mentioned above, the men filled out a mood questionnaire first after they had been emotionally aroused by receiving seven shocks and then again immediately after seeing the film (and before shocking their partner). The conditions did not differ reliably on any of the mood scales completed on the first occasion. Change scores were then computed for each subject on each of the 10 mood scales, and the three film conditions were compared on the 10 change scores. Table 1 summarizes the findings on six of the scales. The first two scales were the only ones of the 10 yielding a significant among-groups effect in the preliminary analyses of variance. The men witnessing the prize-fight scene, surprisingly enough, reported a greater increase in feelings of sadness and unpleasantness than the men shown the non-aggressive track film. As is apparent in the third line, there was no significantly greater increase in experienced anger in the people seeing the fight. Finally, looking at the last three scales, we can note that the film conditions did not differ reliably in reported changes in felt anxiety, tension, and worry. All in all, rather than producing an increased anger or anxiety, the aggressive movie evidently made the men feel somewhat sadder.³

³ It is an interesting question as to why this increased sadness arose. But whatever the reason, this is irrelevant to our present theme. The point remains that the greatest overt aggression was displayed in one of the aggressive film conditions, even though the mood change produced by the film was not one of increased anger.

TABLE 1
MEAN CHANGES IN MOOD FROM BEFORE TO
AFTER FILM

Mood scale	Justified film aggression	Less- justified film aggression	Track film	Among group F ratio
Sad-HAPPY	-0.33 _a	-0.54 _a	0.40 _b	7.3
Pleasant-unpleasant	0.64 _{ab}	0.74 _a	-0.27 _b	4.6
Angry-NOT angry	-0.54 _a	-1.04 _a	-0.40 _b	1.16
Anxious-relaxed	-0.53 _a	-0.40 _a	0.30 _b	2.47
Calm-TENSE	0.33 _a	-0.07 _a	-0.22 _a	1.49
Not worried-worried	0.00 _a	-0.10 _a	0.03 _a	1.63

Note.—Separate analyses of variance were conducted for each scale. In regard to any one scale, cells having a subscript in common are not significantly different, at the .05 level, by Duncan multiple-range test. The capitalized word indicates the direction of a high, positive-change score.

* $p < .05$.

** $p < .01$.

Aggression toward Target Person

Considering the moods existing in the three film conditions, the usual common-sense theory of aggression would undoubtedly not have predicted the results obtained with the shock measure. Over all 90 subjects, the men who had witnessed the aggressive movie subsequently gave somewhat more shocks to their tormenter than did the people seeing the nonaggressive film, even though the former were sadder rather than angrier. But more important than this, the greatest number of shocks were administered by the provoked subjects who had seen *justified film aggression* and whose target had a name (Kirk) associating him with the victim of this observed violence. Kirk generally received somewhat more shocks than did Bob, but only after the subjects had watched a man named "Kirk" being beaten on the movie screen. The only significant difference in this total sample, however, was between the *justified-film-aggression-Kirk* group and the *track-film-Kirk* group.

A secondary analysis of the data was carried out under the assumption that strong anxiety had led to an inhibition of aggressive responses in the most strongly instigated group. Scores on three intercorrelated mood scales, anxious-not anxious, calm-tense, not worried-worried (the three scales shown at the bottom of Table 1), were combined for each subject to yield an anxiety index, and the five most anxious men in each condition were then set aside. We then determined the mean number of shocks given by the less anxious people after the mood scales had been filled out. The results are shown in the bottom half of Table 2. A much clearer picture now emerges. The available target whose name associated him with the victim of justified violence received a

TABLE 2
MEAN NUMBER OF SHOCKS TO CONFEDERATE

Confederate's name	Justified film aggression	Less-justified film aggression	Track film
Total sample ^a			
Kirk	5.87 _a	5.13 _{ab}	4.13 _b
Bob	5.00 _{ab}	4.67 _{ab}	4.60 _{ab}
Omitting 5 most anxious men in each group ^b			
Kirk	6.4 _a	5.0 _b	4.4 _b
Bob	4.8 _b	4.3 _b	4.7 _b

Note.—Cells having a subscript in common are not significantly different, at the .05 level, by Duncan multiple-range test.

^a N = 15 in each group.

^b N = 10 in each group.

significantly greater number of shocks than the target person in any other condition. As we had predicted, the available target's name had a significant effect even when this name-mediated association with the observed aggression was established after the aggression was witnessed.

Questionnaire Measures of Hostility

The four questionnaire measures of hostility toward the confederate gave rise to results essentially similar to those obtained with the electric-shock measure; on three of the four items the strongest disapproval of the confederate was expressed by the men who had witnessed the *justified film aggression* and then rated the confederate named Kirk. This group was tied with the *justified-film-aggression-Bob* condition on the fourth item. Significant condition differences were obtained with only two of the items, however. The results with these two measures are summarized in Tables 3 and 4.

Looking at the findings for the total sample, as shown in the top halves of the tables, we can see that the men provoked by the confederate who

TABLE 3
MEAN EXPRESSED DESIRE TO HAVE CONFEDERATE AS PERSONAL FRIEND

Confederate's name	Justified film aggression	Less-justified film aggression	Track film
Total sample ^a			
Kirk	4.80 _a	4.13 _{bc}	3.67 _c
Bob	4.40 _{ab}	4.13 _{bc}	3.80 _{bc}
Omitting 5 most anxious men in each group ^b			
Kirk	4.70 _a	4.40 _{ab}	3.70 _{bc}
Bob	4.40 _{ab}	4.20 _{abc}	3.60 _c

Note.—Cells having a subscript in common are not significantly different, at the .05 level, by Duncan multiple-range test. A high score indicates an unfavorable attitude.

^a N = 15 in each group.

^b N = 10 in each group.

TABLE 4
MEAN EXPRESSED DESIRE TO HAVE CONFEDERATE AS ROOMMATE

Confederate's name	Justified film aggression	Less-justified film aggression	Track film
Total sample ^a			
Kirk	5.20 _a	4.33 _{ab}	4.20 _b
Bob	4.80 _{ab}	4.53 _{ab}	4.27 _{ab}
Omitting 5 most anxious men in each group			
Kirk	5.20 _a	4.70 _{ab}	4.30 _{ab}
Bob	5.00 _{ab}	4.80 _{ab}	4.00 _b

Note.—Cells having a subscript in common are not significantly different, at the .05 level, by Duncan multiple-range test. A high score indicates an unfavorable attitude.

^a N = 15 in each group.

was associated with the *justified film aggression* subsequently expressed a reliably lower desire to have him as a personal friend (Table 3) and as a roommate (Table 4) than the other subjects also angered by Kirk but who had watched the nonaggressive movie. Although the *justified-film-aggression-Kirk* group was not significantly more hostile to the confederate than was the *justified-film-aggression-Bob* condition, the men in the latter group did not differ reliably from their track-film controls. (This latter difference did become significant in the case of the "personal friend" item when the five most anxious subjects were omitted—bottom half, Table 3—because the *track-film-Bob* mean became somewhat lower.) In general, the questionnaire findings parallel those obtained in the earlier experiment (Geen & Berkowitz, 1966). Instead of there being a cathartic purge of aggressive tendencies in the men giving the greatest number of electric shocks, the strong instigation produced in them by giving them a target associated with the source of their aggressive predisposition evidently led to relatively persistent aggressive response chains; they continued to display the strongest aggression on the next occasion when the aggression was to be expressed verbally rather than physically.

DISCUSSION

Taken together, the findings in the writers' research program point to the importance of considering the available target's stimulus characteristics in any comprehensive analysis of aggression. Particular people are attacked, the senior author has contended, not only because they are safe and visible targets, but also because they have cue properties causing them to elicit aggressive responses from persons who are ready to act aggressively. Supporting this reasoning, available target persons who are associated with the victim of observed violence receive more attacks from angered individuals than do other

possible targets lacking this association. This result is obtained, furthermore, even when the available target's connection with the observed victim is established after the aggressive event is witnessed.

Our results do not unequivocally demonstrate, however, just how aggressiveness cues function. While the previous papers in this series had suggested that target stimuli "pull out" aggressive responses, the possibility still remains that the actions were "pushed out" by a strong, although short-lived, internal arousal state produced by the stimuli. McClelland's (1953) affect conception of motivation takes this type of position. He had proposed that motives were aroused when some cue reintegrated a previously experienced affective state. Hearing a name connected with the just-witnessed aggressive film conceivably could have reintegrated the affect that had been aroused by the film, and this reawakened affect could then have "driven" the attacks upon the target person. The mood changes shown in Table 1 appear to argue against this formulation. The aggressive movie seems to have generated greater sadness than anger, and it is unlikely that this felt sadness had impelled the subsequent aggression. We prefer to suggest that the violent movie had elicited implicit aggressive responses within the men witnessing this film, rather than experienced anger. These implicit responses presumably increased the audience's readiness to act aggressively, enabling the subsequently encountered aggressive cue to evoke strong aggressive actions from the men.

REFERENCES

- BERKOWITZ, L. *Aggression: A social psychological analysis*. New York: McGraw-Hill, 1962.
- BERKOWITZ, L. Aggressive cues in aggressive behavior and hostility catharsis. *Psychological Review*, 1964, 71, 104-122.
- BERKOWITZ, L. The concept of aggressive drive: Some additional considerations. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 2. New York: Academic Press, 1965. Pp. 301-327. (a)
- BERKOWITZ, L. Some aspects of observed aggression. *Journal of Personality and Social Psychology*, 1965, 2, 359-369. (b)
- BERKOWITZ, L., CORWIN, R., & HEIRONIMUS, M. Film violence and subsequent aggressive tendencies. *Public Opinion Quarterly*, 1963, 27, 217-229.
- BERKOWITZ, L., & GEEN, R. G. Film violence and the cue properties of available targets. *Journal of Personality and Social Psychology*, 1966, 3, 525-530.
- BERKOWITZ, L., & RAWLINGS, E. Effects of film violence on inhibitions against subsequent aggression. *Journal of Abnormal and Social Psychology*, 1963, 66, 405-412.
- GEEN, R. G., & BERKOWITZ, L. Name-mediated aggressive cue properties. *Journal of Personality*, 1966, 34, 456-465.
- MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. W., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
- SPENCE, K. W. *Behavior theory and conditioning*. New Haven: Yale University Press, 1956.

(Received March 18, 1966)

Journal of Personality and Social Psychology
1967, Vol. 5, No. 3, 368-372

DISTRACTION AS A FACTOR IN THE ENJOYMENT OF AGGRESSIVE HUMOR¹

HARRY F. GOLLOB

AND

JACOB LEVINE

University of Michigan

*West Haven Veterans Administration
Hospital, Connecticut*

Freud has argued that successful aggressive humor distracts a person so that he is not fully aware of the content of what he is laughing at. If a person focuses his attention on the fact that the humor expresses aggressive impulses, his inhibitions become mobilized and he is then relatively unable to enjoy the humor. Female Ss made humor ratings of cartoons before and after they had their attention focused on the cartoon content by being asked to explain the joke. Although initially highest in humor, high-aggressive cartoons received significantly lower ratings on the posttest than either low-aggressive or nonsense cartoons.

Humor has long been recognized as a social force which plays an important part in our daily

¹ This article is a somewhat expanded version of a paper which was presented at the 1965 American Psychological Association Convention in Chicago.

lives (Levine, in press). Freud (1960) and others have emphasized the close relationship between the dynamics of humor appreciation and defensive processes more generally. Furthermore, Freud's (1960) work provides a useful theoretic

cal framework which allows one to form many eminently testable hypotheses concerning the dynamics of humor appreciation. It is the purpose of this paper to describe an experimental investigation of one such hypothesis.

Freud (1960) has emphasized the importance of distinguishing between tendentious jokes, which serve some sexual or aggressive purpose, and innocent (nontentious) jokes, in which pleasure depends only on the mental activity associated with the joke technique (e.g., incongruity, play on words, representation by the opposite, etc.). Tendentious jokes make possible the expression and partial satisfaction of sexual or hostile impulses in the face of an obstacle which opposes more direct expression of the impulse. Thus, tendentious jokes provide pleasure through both the joke technique (which is common to both innocent and tendentious jokes), and through providing partial satisfaction of impulses which would otherwise remain unexpressed and unsatisfied. When enjoying a tendentious joke, one does not ordinarily know what part of pleasure is due to the joke technique and what part is due to the underlying content. Freud (1960, pp. 150-153) felt that if a person becomes fully aware that a joke expresses unacceptable sexual or hostile impulses, his inhibitions (or awareness of external obstacles) become mobilized and he is then relatively unable to enjoy the joke. Thus Freud argues that in order for tendentious humor to be successful it *must* provide distraction so that the person does not immediately become fully aware of what he is laughing at. Probably the most important source of distraction which the joke provides is that which results from the "innocent" aspects of the joke, that is, the pleasure and laughter which are derived from the joke technique itself.

In the present experiment subjects gave humor ratings of cartoons before and after they had their attention focused on the cartoon content by being asked to explain the joke. Freud's theoretical notions led to the hypothesis that cartoons depicting a high degree of interpersonal aggression, although selected to be the most humorous on the pretest, would on the posttest be rated significantly less funny than either low-aggressive or innocent cartoons.

METHOD

Selection of Cartoons

The cartoons used in this study were of three types: high interpersonal aggression, low interpersonal aggression, or nonsense. Nonsense (innocent) cartoons were defined for judges as "cartoons which contain minimal drive content or anxiety arousing material." From a large pool of cartoons we selected 40 cartoons depicting various degrees of interpersonal aggression and also selected 20 nonsense

cartoons. Seven advanced graduate students in personality and clinical psychology then selected from this smaller pool cartoons whose *only* important theme was that of interpersonal aggression. They placed cartoons which they felt met the criterion for being nonsense cartoons in a separate pile. Each judge then sorted the aggressive cartoons which he had selected into five categories ranging from low- to high-aggressive content.

Six of the seven judges agreed on 30 cartoons whose *only* important theme was that of interpersonal aggression, and also agreed on 13 cartoons which could be considered nonsense (innocent) cartoons. On the basis of the judges' mean aggression ratings for the 30 aggressive cartoons, 8 of the 10 most aggressive and 8 of the 10 least aggressive cartoons were selected for use in the pretest. Eight of the 13 nonsense cartoons were also selected for the pretest.

Study I

Subjects were 14 female college students who participated in the study during a regular session of an introductory psychology course. The course instructor introduced the experimenter to the class and explained that the data were being collected for research purposes only and that neither he nor anyone else in the school would have access to the data. During all phases of the study, work done by subjects was anonymous, and instructions emphasized that the experimenter was interested only in the results of the group as a whole. The 24 cartoons selected for the pretest were presented in booklet form, with each subject receiving the cartoons in a different random order. Subjects were asked to rate each cartoon on an 8-point scale ranging from "very funny" to "not at all funny." On the basis of these pretest humor ratings four cartoons were selected which were high in aggressive content and had a mean humor rating which was slightly higher than the mean rating of the four low-aggressive cartoons and of the four nonsense cartoons which were selected. The 12 cartoons included in the final set are described in Table 1.

Ten days after the pretest, during a regular class session, the same subjects completed the second part of the experiment. For each booklet, the order of presentation of the final set of 12 cartoons was randomly determined. Subjects were asked, for each cartoon in turn, to "state what about the cartoon you think is funny, or is supposed to be funny. Describe as vividly as you can the intended point of the joke. . . ." Immediately after having considered a cartoon in this manner, subjects rated the cartoon on an 8-point scale according to how funny it seemed to them at the time. Subjects were allowed 30 minutes to complete this task. Finally, subjects rated each cartoon according to the intensity of "interpersonal aggression" it portrayed. These aggressiveness ratings overwhelmingly supported our *a priori* division of the cartoons into nonsense (minimal aggression), low-aggression, and high-aggression categories.

Due to the fact that the final set of cartoons was selected on the basis of pretest ratings made by the same group of subjects who took the posttest, the results of Study I were affected to some degree by

TABLE 1

DESCRIPTION OF CARTOONS

High aggression

1. A rifle is rigged to shoot anyone who enters the room. A woman, looking toward the door, calmly and seductively calls out, "It's not locked, honey."

2. On a golf course, one golfer is choking another to death. A third golfer nonchalantly walks over and says to the aggressor, "Pardon me, old man. Your grip's all wrong."

3. A husband and wife are fighting viciously in the presence of two guests. The woman guest leans over and whispers to her husband, "I've heard they don't get along very well."

4. A pipe-smoking gentleman is calmly raking leaves around a tree to which is tied a big woman, presumably his wife. It looks as if he is going to burn her at the stake while doing his fall chores around the estate.

Low aggression

5. A floorwalker accompanied by a large, frowning, formidable-looking woman explains to the adjustment clerk that, "The customer is always right, Benson. Misinformed perhaps, inexact, bull-headed, fickle, ignorant, even abominably stupid, but never wrong."

6. A service-station attendant says, "Oil's dirty," as he casually wipes the oil dipstick on his customer's tie.

7. As they walk past a theater marquee advertising the movie, *Monster from the Slimy Swamp*, a husband annoys his wife by saying, "Look—your mother's name in lights!"

8. A man with a severe hangover slumps miserably in an easy chair while his angry wife plays a recording of him saying, "What! Leave already? Why, it's only three o'clock—the night's young . . . !"

Nonsense

9. A crab with a black and white checkered pattern on its shell is standing by the curb of a swanky hotel. A doorman nonchalantly motions to guide a guest toward the crab. The guest replies, "I'm afraid you misunderstood. . . . I asked for a Checkered cab."

10. Two tipplers coming home from a wild night on the town are gaily staggering up and down walls, as well as back and forth across the sidewalk and street.

11. While his surprised wife watches, a man paints his livingroom by holding a paint roller against the wall as he rides back and forth on a bicycle.

12. A "Robot Adding Machine" equipped with myriad dials and gauges is using an abacus to do its adding.

TABLE 2

MEAN RATINGS OF CARTOON TYPES IN STUDY I

	High aggressive	Low aggressive	Nonsense
Pre	5.55	5.21	5.16
Post	3.29	4.09	4.10
Difference	2.26	1.12	1.06

overall effects of statistical regression opposed the predicted results.²

Study II

Study II was carried out about 3 months after the first study and is basically a replication of Study I. Subjects were 14 female college students fitting the same description as that given for subjects in Study I. The procedure used was essentially the same as in Study I, with the exception that the final set of 12 cartoons was presented in *both* the pre- and posttest. Thus statistical regression did not affect the results of Study II.

TABLE 3

MEAN RATINGS OF CARTOON TYPES IN STUDY II

	High aggressive	Low aggressive	Nonsense
Pre	5.09	5.05	4.36
Post	4.21	4.83	4.48
Difference	.88	.22	-.12

RESULTS

The results of Studies I and II are presented in Tables 2 and 3, respectively. Pre- and posttest humor ratings for the three cartoon types are given in the first two lines of each table, and the pre- minus posttest differences are shown on the third line of the tables. Although there are some large unexplained differences between Tables 2 and 3, the *pattern* of means in both studies is very similar with respect to the hypotheses presently under investigation. Thus, in order to facilitate discussion, the results of the two studies

²Since cartoon selections were made *within* the three types of cartoons, it is statistical regression *within* cartoon types that we must consider. The mean funniness rating of the four high-aggression cartoons finally chosen for use in the posttest was *smaller* than the mean of the eight high-aggression cartoons used in the pretest. The mean rating of the four selected low-aggression and the four selected nonsense cartoons, respectively, was *greater* than the rating for the eight low-aggressive and eight nonsense cartoons used in the pretest. Thus, if only statistical regression within classes were operating, we would, upon retesting, expect the mean funniness rating for the high-aggression cartoons to *increase*, and the ratings for the low-aggression and nonsense cartoons to *decrease*. Thus, the predicted results of this study are *opposed*, rather than *aided*, by the effects of statistical regression within cartoon types.

effects of statistical regression. However, the magnitude of possible effects of such regression was slight, and furthermore the pattern of means for the initial and final sets of cartoons was such that the

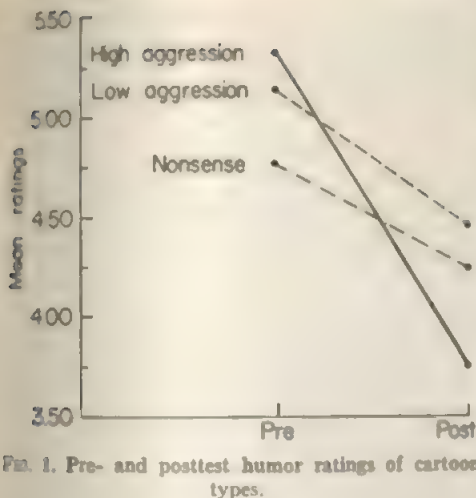


FIG. 1. Pre- and posttest humor ratings of cartoon types.

ies were averaged. The pre- and posttest means so obtained are plotted in Figure 1. Averaging the third line of Tables 2 and 3 yields the mean pre- minus posttest differences of 1.57, .67, and .47 for high-aggression, low-aggression, and nonsense cartoons, respectively. Thus, the greater the aggression content of the cartoons, the greater the decrease in the humor ratings between the pre- and posttest sessions (combined $p < .001^a$). Furthermore, Figure 1 shows clearly that, as hypothesized, the high-aggressive cartoons, although rated funniest on the pretest, received lower funniness ratings on the posttest than did either the low-aggressive (combined $p < .05$) or the nonsense cartoons (combined $p < .05$).

DISCUSSION

Although the results of the present study are consistent with Freud's theory of humor, it is incumbent upon us to briefly discuss some apparently plausible alternative explanations of the data. Perhaps the high-aggressive cartoons differed systematically from the low-aggressive and nonsense cartoons on some dimension other than amount of aggressiveness, and some process occurring along this dimension accounted for the results. We cannot rule out this possibility, but we do regard it as unlikely.

In studies of this type one quite naturally wonders about the role of demand characteristics (Orne, 1962). Subjects in each study were interviewed as a group about what they thought the experimenter was investigating. Although subjects felt the study was concerned in some way

with aggression and humor, no subject seemed even obliquely aware of the possibility that we were interested in seeing whether the decrease in funniness from pre- to posttest was different for different types of cartoon content.

Another problem in interpreting the results of this study is related to the fact that differential pressures to give socially desirable responses may have existed in the pre- and posttest conditions. Thus, subjects' concern about what conclusions the experimenters would draw about them on the basis of their cartoon ratings may have been greater during the posttest than during the pretest. If subjects' "evaluation apprehension" (Rosenberg, 1965) was greater during the posttest it would be expected that the humor ratings of cartoons depicting highly aggressive, socially undesirable behaviors would decrease more than the ratings of cartoons depicting activities with a higher degree of social acceptability. The plausibility of this explanation of the results is weakened by the facts that all work was anonymous, subjects did not communicate with each other during the study, and subjects had no feedback about how others in the class were responding to particular cartoons. Thus, external social pressures to give socially desirable responses were minimal. To the extent that we actually were successful in minimizing subjects' real or imagined perception of such external social pressures, it seems reasonable to interpret the results in terms of intrapsychic defense processes.

It is notable that three of the four high-aggression items and two of the four low-aggression items are concerned with violence between spouses. It is interesting to inquire whether or not the results would hold up with equally aggressive content which is not as close to an area of so much probable conflict for the college girls in this study. Relevant to this point is the fact that the only "nonmarriage" high-aggression cartoon (2, which was rated least funny on the pretest) decreased in its humor rating more than all but one (3) of the "marriage" topic high-aggression cartoons. Furthermore, every high-aggression cartoon decreased more in humor than did any of the low-aggression cartoons.

Since much, if not most, aggressive humor is very obviously aggressive, a question may be raised as to whether or not we are assuming that the subjects in the present study were so naive that they were not aware without prompting that the cartoons were aggressive. It is important to remember that there are varying degrees of awareness. The main idea presented here is that people usually laugh at an aggressive joke or cartoon without thinking vividly and explicitly about the gory details of the aggression. If does happen that a person becomes highly conscious of the fact that the humor expresses

^aThe significance tests used to obtain the "combined p values" presented in this report were based on Stouffer's method (see Mosteller & Bush, 1954, p. 329) for combining the results of two separate investigations. Since a priori hypotheses are being tested, one-tailed p values are reported.

in a sense endorses, an extremely aggressive act, his inhibitions against being too aggressive become mobilized and he then perceives the cartoon as less humorous.

The present study attempted to mobilize subjects' inhibitions surrounding aggression by having them focus their attention on the aggressive content of the cartoons. Shortly following completion of the present study, Singer, Gollob, and Levine (1966) conducted an analogous experiment in which control and inhibition conditions, respectively, were established by exposing male college subjects to artwork depicting either pleasant or brutally sadistic scenes. All subjects then made humor ratings of 12 cartoons depicting various degrees of interpersonal aggression. It was found that the aroused inhibitions interfered specifically with the appreciation of aggressive cartoons, with most interference occurring for cartoons highest in aggressive content. Thus, the study by Singer et al. and the present study complement each other in providing experimental support for the notion that mobilization of strong inhibitions surrounding aggression interferes with the enjoyment of aggressive humor.

In conclusion, although we recognize that alternative explanations may account for the data obtained in this investigation, our positive re-

sults increase our subjective confidence in the heuristic and predictive usefulness of Freud's hypotheses concerning distraction as an important factor in the enjoyment of aggressive humor.

REFERENCES

- FREUD, S. *Jokes and their relation to the unconscious*. (Ed. & trans. by J. Strachey) (Orig. publ. (1905) New York: Norton, 1960.
- LEVINE, J. Humor. In, *International encyclopedia of the social sciences*. New York: Macmillan, in press.
- MOSTELLER, F., & BUSH, R. R. Selected quantitative techniques. In G. Lindsey. (Ed.), *Handbook of social psychology*. Vol. 1. Reading, Mass.: Addison-Wesley, 1954. Pp. 289-334.
- ORNE, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 1962, 17, 776-783.
- ROSENBERG, M. J. When dissonance fails: On eliminating evaluation apprehension from attitude measurement. *Journal of Personality and Social Psychology*, 1965, 1, 28-42.
- SINGER, D. L., GOLLOB, H. F., & LEVINE, J. Inhibitions and the enjoyment of aggressive humor: An experimental investigation. Paper read at Eastern Psychological Association, New York, April 1966.

(Received April 15, 1966)

Journal of Personality and Social Psychology
1967, Vol. 5, No. 3, 372-378

EFFECTS OF CUE EXPLICATION UPON PERSONS MAINTAINING EXTERNAL CONTROL EXPECTANCIES

HERBERT M. LEFCOURT¹

University of Waterloo

3 groups of Ss comprised of an equal number of internal and external control individuals were tested in a level of aspiration situation. Each group received different directions ranging from low to high cue in regard to defining what reinforcements were available in the task. Internal control Ss did not vary between conditions in the incidence of internal control related behavior. External control Ss, however, exhibited a marked increase in internal control related behavior especially as measured by the incidence of patterns in the level of aspiration task. The data supported the hypothesis that external control individuals fail to differentiate situations in regard to reinforcement availability.

In a series of recent investigations (Battle & Rotter, 1963; Gore & Rotter, 1963; Lefcourt &

¹The author wishes to express appreciation to Howard Cappell and Lawrence Lewis without whose help this study could not have been completed.

Ladwig, 1965a; Strickland, 1965) a consistent finding has emerged that oppressed groups such as Negroes in the United States may be characterized as maintaining low expectancies that effort on their own behalf can affect the out-

comes regarding goals they value. The latter two studies have demonstrated that the higher the external control expectancies, the less likely are Negroes to participate in social action designed to ameliorate their conditions. As summarized in two review papers (Lefcourt, 1966; Rotter, 1966) the internal-external control of reinforcement construct has proven to be a useful variable which may help to illuminate certain elusive phenomena such as the noted withdrawal, apathy, and overall avoidance behaviors on the part of oppressed groups, as well as in certain pathological groups (Bialer, 1961; Cromwell, Rosenthal, Shakow, & Kahn, 1961). Briefly, the internal-external control of reinforcement construct derives from social learning theory (Rotter, 1954) and is considered to be a generalized expectancy, operating across a large variety of situations, which relates to whether or not the individual believes that he possesses or lacks power to affect what happens to him. As a general principle, internal control refers to the perception of positive and/or negative events as being a consequence of one's own actions and thereby under personal control; external control refers to the perception of positive and/or negative events as being unrelated to one's own behaviors and therefore beyond personal control. In the two review papers mentioned above, considerable evidence has been reported to the effect that the maintenance of internal or external control expectancies has important ramifications in regard to efforts expended to succeed in affecting the events within one's life space.

Most of the research to date, however, has been concerned with demonstrating the utility of the internal-external control construct. Investigators have focused on predicted group differences or responses to tasks described as more or less controllable. Only one study has focused on attempts to alter external control related behavior (Lefcourt & Ladwig, 1965b). In this investigation Negroes, who had previously been defined as an external control group, were induced to persist in a competitive task with whites when they believed that the skills required in that task were related to an achievement area in which they had already enjoyed some success (jazz). In conclusion, it was argued that behavior on the part of persons maintaining external control expectancies could be altered if new goals could be cognitively linked to whatever prior successes such persons may have had; consequently, they would at least try to make some effort toward accomplishing the new goal.

The present investigation represents a second attempt at altering behavior related to external

control expectancies. The basic assumption of this experiment is that external control expectancies are maintained by their own consequences, apathy and inactivity. If a person believes that there is nothing he can do to alter a faltering marriage, for instance, then he will not be aware of nor ready to take advantage of cues offered by his partner regarding possible avenues open for a rapprochement. To put it simply, the lack of involved struggle in attempting to change an event deprives an individual of the experiences which may be prerequisite to success in affecting that event. In psychoanalytic parlance, a person could be said to fail to develop ego strength because of his overexercised avoidance responses which have prevented direct encounters with life's challenges. It is hypothesized here that one of the major drawbacks of a lack of such involvement is that the external control person is less aware of the cues which could inform him of the probability for success experiences in different situations.

At the same time, there is some evidence to suggest that external control persons may be more suggestible and conforming than internal control persons (Crowne & Liverant, 1963; Odell, 1959). With the assumptions that external control individuals both suffer from inability to recognize cues that might guide them toward success experiences and are more suggestible and conforming than internal control persons, it is anticipated that external control subjects should be more affected by instructions that explicate the purpose of and the methods by which to succeed in given tasks than internal control subjects. The experiment to be described presents a range of directions from least to most explicit for the same level of aspiration task for comparable groups of internal and external control subjects. The overall hypothesis is that behavior in a level of aspiration task will become increasingly indicative of an internal control orientation, for subjects independently assessed as external controls, when instructions increasingly clarify the purposes of the task. In contrast, the internal control subjects will show little variance as a function of such clarified directions.

METHOD

Subjects

All subjects were first-year undergraduates enrolled in the introductory psychology course at the University of Waterloo. Participation in research is required in these courses though the choice of project is optional. The total sample consisted of 132 subjects whose mean age was 19 years.

Procedure

All subjects completed the 23-item Internal-External Control of Reinforcement (I-E) Scale before the major task began. Additional scales were administered to disguise the focus of the investigation. For the purpose of this study, scores on the I-E Scale were divided at the median (7.5) such that all subjects with scores between 0 and 7 were categorized as "internal control," while scores of 8 and above were labeled "external control." For all three conditions to be described there were 22 subjects in each cell, 22 internals and 22 external control subjects within each of the three conditions.

Each subject was then given the Rotter (1954) Level of Aspiration (LOA) Board using Rotter's standard instructions. The experimenters were unaware as to which group, internal or external, the subject belonged in. While these directions communicate how one earns points in the LOA task, they do not emphasize what the meaning or purpose of success or failure might be. Nevertheless, the LOA Board lends itself easily to an achievement interpretation. The difficulty of making high scores, the possibility for improving one's performance with awareness of one's successes and failures can offer a challenge to a person's sense of mastery and competence.

In Condition 1, referred to as the low-cue condition, the task was described as follows: "This game has been used as an amusing diversion with children. We are interested in seeing how adults will perform in this game." No statement was made regarding the importance or meaning of performance. In other words there was no indication of what reinforcements might be contingent upon performance in the LOA task. Consequently, the subject may be described as having no *provided* expectancies regarding the reinforcements for performance unless he supplied the interpretations himself which in the LOA task should have led to an anticipation of achievement reinforcements.

Condition 2 is referred to as the moderate-cue condition. It consists of Rotter's standard directions which explain how to earn points and describe the task as one that is related to motor-control skills. In contrast to Condition 1, an atmosphere of achievement potential is created by the directions. In this condition, the task is described as demanding skill in that there is some importance to succeeding. This condition is seen as providing further cues than Condition 1 as to the meaning of performance. It should be easier in this condition to perceive some possibility of achievement reinforcements if it had not occurred already to the subject in Condition 1.

Condition 3 is the high-cue condition in which Rotter's standard directions are embellished with a greater explication of the demands and purposes of the LOA task:

Actually, this task requires both self-control of your motor movements and insight. You have to try to do well and have to make predictions about

how well you do. In other words, you will be predicting your own performances. Your score will reflect both your control and your insight which we define as accuracy of self-judgment.

In this condition the purpose of good performance is made exceedingly clear. If one does well, then skill in motor control and self-judgment is evident. The subject, therefore, if he believes the directions, should have a clear expectancy that adequacy in the LOA task offers a challenge to one's mastery and skills or, in other words, provides an opportunity for achievement reinforcements.

The task directions become increasingly explicit from Conditions 1 to 3 concerning the meaning of performance. A successful performance in Condition 1 merely indicates that one does well in childhood tasks. The second and third conditions present an increasingly explicit interpretation of success and stress the manner in which one maximizes his success potential.

Hypotheses

It is predicted that external control subjects will exhibit greater responsiveness to the differences in directions noted above than internal control subjects. The latter group should demonstrate more internal control related behavior in the low-cue situation, and somewhat more in the moderate-cue condition than the external control subjects. With high-cue directions, however, it is predicted that the differences between the internal and external control groups will diminish to nonsignificance. In other words, internal control subjects should be relatively consistent across the conditions, maximizing success probabilities to the same degree regardless of directions. External control subjects, on the other hand, are expected to show benefit from the increasingly explicit directions presenting cues on how and why one should succeed.

Four indexes of LOA performance which theoretically and empirically are related to the I-E dimension are employed: (a) the number of shifts, (b) the number of unusual shifts, (c) patterns, and (d) "abnormal" shifting. The number of shifts (changes in the subject's predictions of his scores) in the LOA relates to general stability and self-confidence. High frequencies of shifts characterize individuals who do not use previous experience to establish consistent estimates of their performance. Unusual shifts (up after failure, down after success) likewise suggest failure of the subject to establish reliable estimates of his skill. Unusual shifts tend to indicate dependence on luck or magical, externally controlled factors. Rotter (1954) has described nine patterns characterizing level of aspiration behavior regarded as "points of concentration which can be used as flexible standards for understanding individual cases [pp. 319-322]." These patterns are based upon a combination of three factors: *D* scores, the frequency of shifting, and the occurrence of unusual shifts (up after failure and down after success):

1. *Low positive D score pattern.* This is generally selected a realistic performance. Predictions are higher than past performance with an average number of shifts and an absence of unusual shifts.

2. *Low negative D score pattern.* This pattern is similar to Number 1, with a lower *D* score.

3. *Medium high D score pattern.* This pattern is characterized by higher *D* scores with an average number of shifts and no unusual shifts.

4. *Achievement follower pattern.* Subjects' predictions are constantly changed to closely approximate previous performance.

5. *Step pattern.* The subject never lowers his estimates, consequently experiences a high number of failures.

6. *Very high positive D score patterns.* In this pattern the subject states high goals despite lower performance and may exhibit unusual shifts (primarily raising estimates after failure).

7. *High negative D score patterns.* This pattern is characterized by high negative *D* scores and unusual shifts (primarily down after success).

8. *Rigid pattern.* This pattern is defined by an absence of shifts.

9. *Confused or breakdown pattern.* Here a high number of shifts and frequent unusual shifts in both directions indicate a lack of response to previous experience.

Of these patterns, 1, 2, and 3 require a stable internal control reflecting orientation. The interrater agreement for scoring the nine patterns was 86%; for the pooled internal and external control reflecting patterns agreement reached 95%. The patterns were rated without information regarding the subject group or condition. In previous research, Lefcourt and Ladwig (1965a) used only Patterns 1 and 3 as internal control indexes. However, Pattern 2, like Patterns 1 and 3, requires an average number of shifts, an absence of unusual shifts, and a *D* score that is within "realistic" limits. It is of interest that Pattern 2 was found very rarely among the delinquent samples used by Lefcourt and Ladwig. This more cautious pattern seems more prominent in college samples. In Pattern 1, *D* scores usually range from 0 to 3.0; in Pattern 2, the range is from -2.5 to .5; in Pattern 3, the range is from 3.0 to 6.0. These patterns demonstrate the task involvement and realistic self-appraisal characteristic of persons maintaining internal control expectancies. These three indexes have been used with some success in studies of the I-E dimension (Lefcourt & Ladwig, 1965a; Simmons, 1959). In addition, an abnormal shifting measure is used. Where the number-of-shifts index allows for a comparison between high and low amounts of shifting, the abnormal shifting index allows an evaluation of extremes of shifting which indicate predictions that are noncontingent upon experience. This index has been used before. A sample of external control subjects exhibited more abnormal shifting than internal control subjects (Lefcourt, 1963).

The specific predictions are that as directions be-

come more negative, the number of shifts will increase, and the number of unusual shifts will increase. In the first pattern, subjects will raise their estimates after failure, and in the second pattern, subjects will lower their estimates after success. In the third pattern, subjects will maintain a high level of performance, and in the fourth pattern, subjects will maintain a low level of performance. In the fifth pattern, subjects will maintain a high level of performance, and in the sixth pattern, subjects will maintain a low level of performance. In the seventh pattern, subjects will maintain a high level of performance, and in the eighth pattern, subjects will maintain a low level of performance. In the ninth pattern, subjects will maintain a high level of performance, and in the tenth pattern, subjects will maintain a low level of performance.

RESULTS

As indicated in Table 1, some of the predicted differences in LOA performance were obtained

TABLE 1
LEVEL OF ABNORMAL SHIFTS

Condition	Low	Medium	High	Significance Test
No shifts				
Internal				
<i>M</i>	11.45	10.82	10.32	$F_{(2,10)} = 1$
<i>SD</i>	4.45	5.86	3.43	$F_{(2,10)} = 26.00$
External				
<i>M</i>	12.45	11.77	9.00	$F_{(2,10)} = 1.03$
<i>SD</i>	4.05	4.53	3.23	
No unusual shifts				
Internal				
<i>M</i>	1.18	1.27	1.00	$F_{(2,10)} = 1$
<i>SD</i>	1.82	1.77	1.86	$F_{(2,10)} = 1.03$
External				
<i>M</i>	1.41	1.18	.50	$F_{(2,10)} = 1$
<i>SD</i>	1.68	2.18	.64	
Proportions of Ss with 1 or more unusual shifts				
Internal	45	55	45	<i>ns</i>
External	55	45	50	<i>ns</i>
Proportions of Ss with 2 or more unusual shifts				
Internal	27	27	18	<i>ns</i>
External	36	18	09	$\chi^2 = 5.05^*$
Proportions of Ss with Patterns 1-2-3				
Internal	.50	.64	.55	<i>ns</i>
External	.18	.46	.91	$\chi^2 = 23.78^{***}$
Proportions of Ss with abnormal shifting				
Internal	.50	.36	.36	$\chi^2 = 3.09$, <i>ns</i>
External	.64	.59	.09	$\chi^2 = 16.36^{***}$

* $p < .10$.

** $p < .05$.

*** $p < .001$.

It was predicted that as the instructions change from low- to high-cue conditions, there would be a decrease in the amount of shifting. An overall analysis of variance supported this prediction, $F = 3.26$, $p < .05$ ($df = 2, 126$). However, no interaction with the I-E Scale was obtained. The main effect between conditions was largely accounted for by the difference between the low- and high-cue conditions ($t = 2.74$, $p < .01$). The high-cue condition differed from the moderate-cue condition, but not strongly so ($t = 1.78$, $p < .10$). No difference was found between the low- and moderate-cue conditions in the number of shifts ($t < 1$).

The number of unusual shifts revealed little or no differences between conditions, or for the internal versus external control groups. Both parametric and nonparametric analyses were performed on this measure. The nonparametric analyses were undertaken because of the commonly skewed distribution, with the unusual shift measure. Subjects making no unusual shifts were compared with those making one or more unusual shifts. As is obvious from Table 1, there was little difference between conditions and subjects on this index. A second comparison of those making zero or one versus those making two or more unusual shifts was also made. This was done to minimize the possibility that subjects were included in the unusual shift group who made an occasional unusual shift for justifiable reasons, that is, when a previous failure was just short of the stated goal. In this latter analysis no differences were obtained among conditions for the internal control group. However, among the external control subjects, a chi-square of 5.05, $p < .10$ was obtained. This difference among conditions was largely accounted for by the difference between external control subjects in the low- and high-cue conditions ($t = 2.31$, $p < .05$).

While the proportion of subjects making one or more unusual shifts was roughly equivalent to that proportion found in previous internal control samples (48% in Lefcourt & Ladwig, 1965a), the two or more unusual shift index demonstrates that the external control subjects in this experiment behaved in a far less external control manner than their counterparts in previous research that employed delinquents as the subject population. (Fifty-two percent of the external control sample in the Lefcourt & Ladwig investigation made two or more unusual shifts.)

In contrast to the less than definitive results obtained with the previous measures, the results obtained with respect to the incidence of pat-

terns are dramatic. While the internal control subjects appear fairly constant across conditions, the incidence of Patterns 1, 2, and 3 varying from 50% to 64%, the external control subjects change beyond expectation. Where the $\chi^2 = .86$ between conditions for the internal control subjects, $\chi^2 = 23.78$, $p < .001$ for the external control subjects between conditions. It had been initially predicted that internal control subjects would exhibit a greater incidence of internal control related behavior in the low-cue condition, somewhat more in the moderate-cue condition, and an equivalent incidence with external control subjects in the high-cue condition. In the low-cue condition internal control subjects more often performed Patterns 1, 2, and 3 than external control subjects ($t = 2.25$, $p < .05$). In the moderate-cue condition internal control subjects again exceeded external control subjects in the incidence of the internal control patterns. However, the difference failed to reach statistical significance ($t = 1.21$, $p > .10$). However, in the high-cue condition, the external control subjects did not simply equal the internal control subjects, but exceeded them significantly ($t = 2.69$, $p < .02$). In a comparison between conditions within the internal and external control groups, no differences were found for the internal control subjects. For the external control subjects, however, the differences between low- and moderate-cue conditions produced a $t = 2.00$, $p < .10$; between low- and high-cue conditions, $t = 4.90$, $p < .001$; and between the moderate- and high-cue conditions, $t = 3.38$, $p < .01$. Needless to say, the external control individuals exhibited a remarkable change between conditions as represented in the pattern measure.

Since the number of shifts and the number of unusual shifts failed to demonstrate the excessive change noted in the pattern incidence, it was assumed that the frequency of abnormal shifting (too frequent, more than 12; too infrequent shifting, less than 3) might better help to account for the pattern changes than had the other shifting indexes. As noted in Table 1, the frequency of abnormal shifting did not vary significantly for internal control subjects ($\chi^2 = 3.09$, $p > .10$). However, external control subjects exhibited a decided decrease in abnormal frequency of shifting in the high-cue condition, such that the $\chi^2 = 16.36$, $p < .001$. As is obvious from the data the high-cue condition produced almost the entire change noted in this large chi-square.

In general, the data regarding patterns and abnormal frequency of shifting support the hypothesis strongly, while the number of shifts and number of unusual shifts lend only weak sup-

port for the differential predictions made for internal and external control subjects. While the number of shifts did vary significantly with conditions, it failed to do so differentially for the internal and external control groups. The unusual shift, however, appears to be sufficiently rare that within a student population there seems to be little chance for significance to develop.

As has been predicted, the external control subjects were the most affected by changing conditions, while internal control groups showed little or nonsignificant changes in response to the different conditions.

DISCUSSION

The consistent findings that persons who maintain external control expectancies tend to withdraw from challenges and avoid involvement should suggest possible avenues for altering the incidence of avoidance and withdrawal behaviors. In this investigation one possible approach is explored. It has been asserted that a person who believes that desired outcomes are not within his personal control will be less vigilant or perceptive regarding opportunities for obtaining desired reinforcements. In this experiment, the directions to the LOA task have been varied so as to provide a gradually increasing external definition of the situation. In other words, it becomes progressively clear what type of reinforcements are available in the LOA task.² That persons maintaining external control expectancies alter their LOA performance so radically with an explication of cues regarding reinforcement possibilities raises the issue that a lack of goal-striving behavior may be more adequately predicted on the basis of cognitive and perceptual type deficiencies than from a lack of motivation for those goals. It is conjectured here that the individual with external control expectancies does not adequately search for reinforcement opportunities. It is possible that he fails to maintain the kind of internal dialogue that would facilitate the cognitive sorting and categorizing of the situations so that the opportunities for reinforcement in different situations would be more self-evident. The more explicit directions in the high-cue condition, however, allow for the missing cognitive link for external control subjects. The findings suggest that the provision of an explicit interpre-

tation as to the meaning of behavior in given situations can rectify some of the avoidance behavior of external control individuals. It also suggests that external control individuals may be extremely ready to benefit from external direction, more so than the internal control person who, in a sense, already has decided what reinforcements are available. The very tractability of the external control person may eventually prove to be his greatest asset for overcoming the self-defeating avoidance behavior for which he is noted. The question may be raised, however, as to whether the external control individual can ever function independently of such external direction. Further attention needs to be given to the problem of internalization or the development of the generalized expectancy of internal control.

It can be argued that the changes in directions merely altered the value of the reinforcements in the situation, and that change in behavior, therefore, simply mirrors an increasing need arousal. It should be noted, however, that the LOA task was apparently not seen differently across conditions by internal control individuals. It is assumed here that a fair proportion of university students may be described as generally motivated to succeed in achievement tasks. The difference between the internal and external control students is obtained not because of widely differing degrees of achievement motivation, but in their readiness to perceive or construe a given situation as making achievement reinforcements available. The internal control individual is quicker to perceive the challenge to mastery or achievement reinforcements inherent in the LOA task. The external control individual, on the other hand, needs to be informed that achievement reinforcements are indeed available. The fact that both internal and external control subjects may be characterized as potentially achievement motivated is evident in the reliably low correlations between the I-E Scale and need for achievement measures: $r = -.27$ (Lichtman & Julian, 1964), $r = -.25$ (Odell, 1959). In addition, an increase in achievement behavior with increased expectancies for success has been found for a group characterized as maintaining external control expectancies (Lefcourt & Ladwig, 1965b). It can be concluded, then, that goal-directed behavior can be predicted not only through measurement of motivation, but also of the expectancy that successful performance in a given task will eventuate in valued reinforcements. The problem of increasing goal-striving behavior in a person maintaining external control expectancies lies in helping him become

² Again, it should be noted that the LOA task is easily construed as an achievement task such that a person finding an LOA Board with no experimenter present may become involved in testing out his skills. The writer has often had to discourage patients and colleagues from continuous voluntary performance.

aware of the availability of reinforcements, and perhaps, of the methods for maximizing his chances of succeeding in given tasks.

One shortcoming of the study derives from the type of data collected. Frequency data do not allow one to definitively speak about the effects of conditions. Future work needs to be done using data of a more interval scale nature with repeated-measurement designs so as to ascertain changes in specific individuals as opposed to the independent sample design utilized in this investigation.

REFERENCES

- BATTLE, E., & ROTTER, J. B. Children's feelings of personal control as related to social class and ethnic group. *Journal of Personality*, 1963, 31, 482-490.
- BIALER, I. Conceptualization of success and failure in mentally retarded and normal children. *Journal of Personality*, 1961, 29, 303-320.
- CROMWELL, R., ROSENTHAL, D., SHAKOW, D., & KAHN, T. Reaction time, locus of control, choice behavior and descriptions of parental behavior in schizophrenic and normal subjects. *Journal of Personality*, 1961, 29, 363-380.
- CROWNE, D. P., & LIVERANT, S. Conformity under varying conditions of personal commitment. *Journal of Abnormal and Social Psychology*, 1963, 66, 547-555.
- GORE, P. M., & ROTTER, J. B. A personality correlate of social action. *Journal of Personality*, 1963, 31, 58-64.
- LEFCOURT, H. M. Internal versus external control of reinforcement: A review. *Psychological Bulletin*, 1966, 65, 206-220.
- LEFCOURT, H. M. Some empirical correlates of Negro identity. Unpublished doctoral dissertation, Ohio State University, 1963.
- LEFCOURT, H. M., & LADWIG, G. W. The American Negro: A problem in expectancies. *Journal of Personality and Social Psychology*, 1965, 1, 377-380. (a)
- LEFCOURT, H. M., & LADWIG, G. W. The effect of reference group upon Negroes task persistence in a biracial competitive game. *Journal of Personality and Social Psychology*, 1965, 1, 668-671. (b)
- LICHTMAN, C. M., & JULIAN, J. W. Internal versus external control of reinforcement as a determinant of preferred strategy on a behavioral task. Paper read at Midwestern Psychological Association, St. Louis, April-May 1964.
- ODELL, M. Personality correlates of independence and conformity. Unpublished master's thesis, Ohio State University, 1959.
- ROTTER, J. B. *Social learning and clinical psychology*. Englewood Cliffs, N. J.: Prentice-Hall, 1954.
- ROTTER, J. B. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 1966, 80(1, Whole No. 609).
- SIMMONS, W. Personality correlates of the James-Phares scale. Unpublished master's thesis, Ohio State University, 1959.
- STRICKLAND, B. The prediction of social action from a dimension of internal-external control. *Journal of Social Psychology*, 1965, 66, 353-358.

(Received April 22, 1966)

ERRATA

In the article by Ronald Taft in the May 1966 issue, the third sentence in the first paragraph of the first column on page 602 should read: "For the purpose of this study, differential accuracy may be taken as: total accuracy - (stereotype accuracy + accuracy due to assimilative projection)."

In the article by Esther S. Battle in the December 1966 issue, the second sentence in the first paragraph of the second column on page 639 should read: "Holding a high minimal goal then seems to be a debilitating factor, for this student performs no more competently than the child who holds low expectations and low minimal goals ($t = 1.35$, $p > .05$)."

SUSPICION OF DECEPTION: IMPLICATIONS FOR CONFORMITY RESEARCH¹

LAWRENCE J. STRICKER, SAMUEL MESSICK

Center for Psychological Studies, Educational Testing Service, Princeton, New Jersey

AND DOUGLAS N. JACKSON

University of Western Ontario

Ss' suspicions were appraised about 2 conformity procedures: a simulated group version of the Asch situation and questionnaires with fictitious norms. Many Ss suspected that the purpose of both procedures was to determine whether their responses would be influenced by others that they did not hear spontaneous responses by others in the simulated group, and that the normative answers reported on the questionnaires were incorrect. Boys were generally more suspicious than girls. The suspicion variables were highly interrelated for both sexes and, in general, negatively related to conformity measures. Correlates of the suspicion variables included acquiescence and social desirability response styles, acquiescence, self-esteem, and intelligence.

Deception is a widely used tactic in current psychological research. The deceptions take a variety of forms and are frequently involved in the explanation of the purpose of the study, in the feedback given to the subjects, and in the descriptions of the other participants and of the mechanics of the experiment. Certain kinds of studies, particularly in the areas of personality and social psychology, routinely use some form of deception. Indeed, it is difficult to imagine how some phenomena could be investigated otherwise.

The study of conformity in particular is steeped in deception. Since the beginning of this line of research at the turn of the century (Bridges, 1914; Brown, 1916; Münsterberg, 1914; Terman, 1904), investigators have been concerned about the plausi-

bility of the deceptions that they employed and have sometimes attempted to assess their effectiveness. However, in common with deception studies in other areas, such appraisals are the exception rather than the rule, and those that are made are often rather superficial (Stricker, 1966). In view of the important role of this tactic in conformity research, a systematic evaluation of its usefulness is imperative.

The purpose of the present study was to assess the extent of subjects' suspicions about deceptions and to determine some of the characteristics of suspicious subjects, using data that were routinely obtained in the course of a large study of conformity.

METHOD

Subjects

The sample consisted of 190 paid volunteers, 101 boys and 89 girls, who were either in the eleventh or twelfth grade or had just graduated. All of the girls were from Princeton, New Jersey, while 62 of the boys were from Princeton, and 39 were from Pennington, New Jersey, a similar neighboring town.

Procedures

There were three data-gathering sessions. The first and third involved group-administered procedures, and the second was devoted to simulated-group laboratory procedures. At the first, several instruments were administered in the following order: a self-report personality inventory containing response style and content scales; the Information-True test,

¹ This study was supported in part by the National Institute of Mental Health, under Research Grants M-2878 and M-4186, and in part by the National Institute of Child Health and Human Development, under Research Grant 1 P01 HD 01762-01. Portions of this study were presented at the meetings of the American Psychological Association, New York, September, 1966.

Thanks are due Robert Fried for designing, constructing, and installing the apparatus and acting as an experimenter; Gloria DuBois for acting as an experimenter; Joseph Notterman for providing laboratory space in the Department of Psychology, Princeton University; and Henrietta Gallagher for supervising the statistical analyses.

a measure of acquiescence; and tests and questionnaires containing the items used in the conformity measures. All were administered with standard test-taking instructions. At the second session, subjects appeared in groups of five for the Olmstead and Blake (1955) simulated-group version of the Asch (1956) situation. At the end of this session, subjects completed an open-ended questionnaire concerning their perceptions of the study's purpose and the behavior of the other subjects. At the last session, held 14-16 days after the initial one for the Princeton subjects and 5 days after the first session for the Pennington subjects, the second version of two questionnaire measures of conformity was administered, this time containing the purported average answers for the group. These questionnaires were followed by another open-ended questionnaire similar to the first. At the end of this session the deceptions and purpose of the study were fully described to the subjects.

Personality Measures

The personality inventory administered at the first session contained these scales:²

1. Fulkerson's (1958) Acquiescence scale.
2. Messick's (1962) Personality Research Inventory (PRI) Acquiescence scale—combines PRI-1 and PRI-2 scales.
3. Bass' (1956) Social Acquiescence scale.
4. Couch and Keniston's (1960) Agreement Response Scale (ARS).
5. Clayton and Jackson's (1961) Acquiescence to Tentatively-Worded F Scale Items (PF Acquiescence)—six authoritarian and six nonauthoritarian items were used, none overlapping with the AF Acquiescence scale items—the score being the number of "true" responses.
6. Clayton and Jackson's (1961) Acquiescence to Extremely-Worded F Scale Items (AF Acquiescence)—six authoritarian and six nonauthoritarian items were used, none overlapping with the PF Acquiescence scale items—the score being the number of "true" responses.
7. Attitude Acquiescence scale—a 40-item scale containing four items, with social desirability (SD) ratings in the neutral range, from each of 10 content areas,³ the score being the number of "true" responses.

² The following scales together with their scoring keys have been deposited with the American Documentation Institute: PF Acquiescence, AF Acquiescence, PF Authoritarianism, AF Authoritarianism, Attitude Acquiescence, Attitude SD, Ascendancy, Self-Esteem, Extroversion, Neuroticism, and Independence-Yielding. Order Document No. 9269 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance \$2.50 for microfilm or \$6.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

³ The 10 areas are capital punishment, education, militarism, minorities, nationalism, newspapers, rapid

8. Edwards' (1957) SD scale.
 9. Stricker's (1963) SD scale.
 10. Messick's (1962) PRI SD scale—combines SD-1 and SD-2 scales.
 11. Wiggins' (1959) Sd scale.
 12. Marlowe-Crowne (Crowne & Marlowe, 1960) SD scale.
 13. Attitude SD scale—a 40-item scale containing four items—two with extremely high SD ratings and two with extremely low ratings—from each of the same 10 content areas as the Attitude Acquiescence scale, the score being the number of socially desirable responses.
 14. Ascendancy scale—16 items from the Ascendancy scale of the Guilford-Martin (1943) Inventory of Factors GAMIN, not duplicating or similar to the items used on the other Guilford scales in this study, were randomly selected and revised by (a) reversing, on a random basis, enough items so that each scale was balanced, and (b) rewriting items so that they were in the first person, clear, and easy to read.
 15. Self-Esteem scale—22 items from the Self-Confidence scale of the Guilford-Martin (1943) Inventory of Factors GAMIN were selected and revised in the same way as the Ascendancy scale.
 16. Extroversion scale—22 items from the Rathymia scale of Guilford's (1940) An Inventory of Factors STDCR were selected and revised in the same way as the Ascendancy scale.
 17. Neuroticism scale—20 items from the Cycloid Disposition scale of Guilford's (1940) An Inventory of Factors STDCR were selected and revised in the same way as the Ascendancy scale.
 18. Independence-Yielding scale—20 of the items that Barron (1953) and Crutchfield (1955) found were associated with conformity, and that did not overlap with items on the other scales in this study, were selected and revised so that half were keyed "true" and half "false."
 19. Clayton and Jackson's (1961) Authoritarianism on Tentatively-Worded F Scale Items (PF Authoritarianism)—uses the same items that appear on the PF Acquiescence scale, the score being the number of authoritarian responses.
 20. Clayton and Jackson's (1961) Authoritarianism on Extremely-Worded F Scale Items (AF Authoritarianism)—uses the same items that appear on the AF Acquiescence scale, the score being the number of authoritarian responses.
- The inventory was followed by Gage, Leavitt, and Stone's (1957) Information-True test, as modified by Stricker (1963).

Intelligence

Deviation IQ scores from the Henmon-Nelson Tests of Mental Ability (Lamke & Nelson, 1957), administered in the ninth or tenth grades, were obtained from school records for 146 of the subjects, and the scores on this test, administered in the eighth

changes versus gradualism, religious conventionalism, rural versus urban living, and science.

or eleventh grades, were available for 16 of the remaining subjects. The scores for 19 other subjects were estimated from their Verbal and Mathematics subtest scores on the Preliminary Scholastic Aptitude Test (PSAT: College Entrance Examination Board, 1965a, 1965b), using a multiple-regression equation based on 99 subjects who had both Henmon-Nelson and PSAT scores. Mean scores for their sex were assigned to nine subjects who had neither Henmon-Nelson nor PSAT scores.

Simulated-Group Conformity Measures

Two separate tasks were used in the simulated-group procedure: (a) counting metronome clicks, and (b) reporting agreement or disagreement with a group of heterogeneous attitude items.⁴ The same items, and in the same sequence, were used in the initial group-testing session and in the subsequent simulated-group session. At the initial session, 28 series of clicks were presented over a loudspeaker, and subjects recorded the number of clicks that they heard. The clicks had a rate of 160 per minute. At the same session, the 20 attitude statements used for the simulated-group procedure, together with items used in the questionnaire procedure, were administered in an Opinion Questionnaire with standard instructions to indicate agreement or disagreement with each statement, using a 9-point Likert scale.

At the simulated-group session, the five subjects—all of the same sex—were conducted to separate rooms, told how to use the microphone in the room, and fitted with earphones. A male experimenter was used with boys and a female experimenter with girls. Subsequently, the subjects were given tape-recorded instructions that the study concerned their ability to use a communications network, and that they would be called on to report the number of clicks they heard and their agreement or disagreement with the attitude statements. Each subject was told that he would be called on after he heard the reports of the other four subjects who were present. Actually, each subject heard the tape-recorded responses of confederates—boys heard boys' voices, girls heard girls' voices. On some trials each confederate made the same response, on others they varied. On some trials the median of the reported responses was correct for the clicks or was the median response to the attitude statements in a similar group of subjects. On the critical items, which were scored for conformity, the confederates' responses—all gave the same answer—systematically differed from the correct number of clicks or the median response on the attitude items. A conformity score was derived for both the clicks and the atti-

tude statements. It was the number of critical items on which the subject diverged in response between the initial testing session and the simulated-group session in the direction of the confederates' responses.

Twenty of the 28 click trials were critical. The number of clicks per trial ranged from 17 to 32 for the critical items and from 19 to 31 for the other items. The confederates' responses were too low on half of the critical items and too high on the other half. Their responses differed from the correct response by one click or trials of 4 or fewer clicks and by one or two clicks on trials of 25 or more clicks.

Fourteen of the 20 attitude statements were critical. On half of the critical items, the median response by high school students in Cohasset, Massachusetts (a community similar to Princeton and Pennington), was Strongly Agree or Moderately Agree, and the confederates' responses were Extremely Disagree, Strongly Disagree, or Moderately Disagree. On the other half of the critical items, the median response was Extremely Disagree, Strongly Disagree, or Moderately Disagree, and the confederates' responses were Extremely Agree, Strongly Agree, or Moderately Agree. The difference between the median response and the confederates' responses was roughly two or three standard deviations.

Questionnaire Conformity Measures

There were two questionnaire measures of conformity.⁵ One, the Test of Accuracy in Estimating Probabilities, used 45 items from a questionnaire employed by Wallach and Kogan (1959) and adapted from one devised by Brim (1955). The items were of the form: "The chances that an American astronaut will reach the moon before 1970 are about _____ in 100." The subject was to respond with an answer ranging from 0 to 100. The other, the Opinion Questionnaire, contained 39 heterogeneous attitude statements. This questionnaire was modeled after one constructed by Hastorf and Piper (1951) and subsequently modified by Jackson (1964). The subject was to indicate his agreement on a 9-point Likert scale. Both questionnaires were administered with the same items at both the initial and second group-testing sessions. At the initial session, the questionnaires were administered with standard test-taking instructions. At the second session, these instructions were preceded by special ones which indicated that the questionnaires were being readministered in order to determine their stability, and that the average answer made to each item by the subjects at the initial testing session was being reported to satisfy any interest that the subjects might have. The reported answers were fictitious. They were the average responses for similar groups of subjects on some items, but on the critical items they were systematically different from these actual average responses. On

⁴ Transcripts containing the verbatim instructions to subjects, the number of metronome clicks used in each trial, and the attitude statements, together with the confederates' responses and the scoring keys, have been deposited with the American Documentation Institute. See Footnote 2 for ordering information.

⁵ These questionnaires together with the scoring keys have been deposited with the American Documentation Institute. See Footnote 2 for ordering information.

each questionnaire, the conformity score was the number of critical items on which the subject changed his responses between the two group-testing sessions in the direction of the reported average.

Twenty-four of the 45 items on the Test of Accuracy in Estimating Probabilities were critical. On the noncritical items, the reported answer was the mean response for summer students at Temple University. The critical items consisted of 12 pairs of items matched on mean and standard deviation for the Temple University students; the mean was above 50 for half the pairs. The reported answer was 30 points (about one and one-half standard deviations) above the actual mean for one item in the pair and 30 points below the actual mean for the other item in the pair. Twenty of the 39 items on the Opinion Questionnaire were critical. On the noncritical items, the answer reported was the actual median response for Cohasset high school students; on half the critical items, their median response was Strongly Agree or Moderately Agree, and the reported answer was Extremely Disagree, Strongly Disagree, or Moderately Disagree. On the other half, the actual median for Cohasset students was Extremely Disagree, Strongly Disagree, or Moderately Disagree, and the reported answer was Extremely Agree, Strongly Agree, and Moderately Agree. The difference between the median and the reported answer was about two or three standard deviations.

Suspicion of Deception Measures

On the basis of the replies on the open-ended questionnaires, two research assistants, working independently, classified the subjects as suspicious, unsuspicious, or indeterminate about the purpose of each of the two kinds of conformity procedures and of the methods that they employed.⁶ A subject was considered to be suspicious about the purpose of the study if he thought it was to determine if his responses were influenced by knowledge of the other subjects' responses. He was classified as suspicious about the method if he believed that the responses attributed to the other subjects were not spontaneous or were intentionally incorrect. The judges rated 173 questionnaires about the simulated group and 155 about the questionnaire procedures. Neither judge was familiar with the details of the study, its purpose, or the hypotheses involved. The order in which the ratings were made was counterbalanced. One judge made ratings of the purpose of the simulated group and the purpose of the questionnaires, followed in a week by ratings of the method of the questionnaires and the method of the simulated group; the other judge made ratings of the method of the simulated group and the method of the questionnaires and a week later, the purpose of the questionnaires and the purpose of the simulated group.

⁶ The questionnaires and the judges' rating instructions have been deposited with the American Documentation Institute. See Footnote 2 for ordering information.

The agreement between the judges on the four ratings was as follows: purpose—simulated group, 91%; method—simulated group, 95%; purpose—questionnaire, 91%; and method—questionnaire, 93%. Ratings on which the two judges disagreed were resolved by obtaining an independent rating from a third judge. If this rating was the same as one of the original ratings, it was assigned to the subject; if it was different from both of the original ratings, the subject was given an "indeterminate" rating.

RESULTS

Illustrative Protocols

As an illustration of the flavor of the subjects' responses on the questionnaires used in assessing their suspicion, extracts from randomly selected protocols are reproduced below.

Suspicious about the purpose of the simulated group. A boy wrote that the purpose of the study was "to determine how someone's ideas and answers are influenced and/or changed after they hear the answers or opinions of others."

Unsuspicious about the purpose of the simulated group. Another boy replied "I'm not really sure" of the purpose of the study.

Suspicious about the method of the simulated group. A girl asserted that the laboratory session was "fixed!," that the other four students' ". . . questions were different—or they were reading answers from a sheet you gave them," and that they gave the answers they did "to through [sic] me off."

Unsuspicious about the method of the simulated group. Another girl replied that "Aside from irksome differences of opinion [the other students] behaved well. . . . I on the whole agreed with the first test answer but seldom if at all on the second which bothered me." In commenting on the correctness or appropriateness of their answers, she wrote, "That is for them to decide. I can only say for my part they were correct."

Suspicious about the purpose of the questionnaires. A girl wrote, "I think [the purpose of the study was] to see if we would conform with the ridiculous average answers and if they influenced us any. The latter part I think was to see if we would stick to our former ideas or if we really weren't sure and would change them."

Unsuspecting about the purpose of the questionnaires. Another girl indicated that she thought its purpose was "to make a mockery of the United States" and that the questionnaires were given twice "to see if the persons answers differed afterward."

Suspicious about the method of the questionnaire. A boy explained why he thought the answers that his group gave were not correct or appropriate: "If these are really their answers were theirs [sic] then I think they were sometimes stupid. If it was like the last test then you made them up."

Unsuspecting about the method of the questionnaire. Another boy wrote that the questionnaires were "fun and interesting," that the other students felt "about the same," and that he agreed "not to [sic] much" with their answers, but thought their answers were "correct."

Percentage of Subjects Classified as Suspicious

The percentage of subjects classified as suspicious, unsuspecting, or indeterminate on each rating appears in Table 1. The percentage classified as suspicious ranged from 9% (for girls' method suspicion about the simulated group) to 61% (for boys' purpose suspicion about the questionnaires). Subjects

TABLE 1

PERCENTAGE OF SUBJECTS CLASSIFIED AS SUSPICIOUS, UNSUSPICIOUS, OR INDETERMINATE ABOUT THE PURPOSES AND METHODS OF THE SIMULATED GROUP AND QUESTIONNAIRES

Classification	Simulated group			
	Purpose		Method	
	Boys	Girls	Boys	Girls
Suspicious	55.7	38.8	42.0	9.4
Unsuspecting	42.0	58.8	56.8	85.9
Indeterminate	2.3	2.3	1.1	4.7
	Questionnaires			
	Boys	Girls	Boys	Girls
Suspicious	61.2	48.0	35.0	18.7
Unsuspecting	36.2	52.0	62.5	80.0
Indeterminate	2.5	0.0	2.5	1.3

Note.—A total of 88 boys and 85 girls were run in the simulated group, and 80 boys and 75 girls completed the questionnaires. The number of subjects varies due to missing data.

with indeterminate ratings have been eliminated from subsequent analyses.

Differences between the sexes in the frequencies classified as suspicious and unsuspecting on each rating were assessed by a two-tailed chi-square test, corrected for continuity. The boys were significantly ($p < .05$) more suspicious than the girls about the purpose of the simulated group ($\chi^2 = 4.35$) and its method ($\chi^2 = 21.17$), as well as the method of the questionnaires ($\chi^2 = 4.66$). The boys and girls did not differ significantly ($p > .05$) in their suspicion about the purpose of the questionnaire ($\chi^2 = 2.83$).

Differences between the ratings in the frequencies classified as suspicious and unsuspecting were assessed by a two-tailed chi-square test for dependent groups, corrected for continuity. There were no significant differences for the boys ($p > .05$). They were equally suspicious of the purpose of the two procedures ($\chi^2 = 3.44$), of their methods ($\chi^2 = 3.44$), of the purpose and the method of the simulated group ($\chi^2 = .00$), and of the purpose and the method of the questionnaires ($\chi^2 = .02$). The girls did not differ significantly ($p > .05$) in their suspicion of the purpose of the two procedures ($\chi^2 = 1.02$), but they were significantly ($p < .01$) more suspicious about the method of the questionnaires than the method of the simulated group ($\chi^2 = 39.05$), and more suspicious about the purpose than the method of both the simulated group ($\chi^2 = 29.25$) and the questionnaires ($\chi^2 = 10.58$).

The percentage of boys and girls from Princeton classified as suspicious, unsuspecting, or indeterminate about the purpose and method of the simulated group for each of the 2 weeks that the simulated group was run appears in Table 2. (Pennington boys were excluded from this analysis because they went through the procedure during the same week.) The differences between the 2 weeks in the frequencies classified as suspicious or not suspicious were appraised by a two-tailed chi-square test corrected for continuity. Where the expected frequencies were too small for chi-squares, the Fisher exact probability test was substituted, also using two-tailed probabilities. Neither sex differed significantly ($p > .05$) between the 2 weeks

TABLE 2

PERCENTAGE OF SUBJECTS CLASSIFIED AS SUSPICIOUS, UNSUSPICIOUS, OR INDETERMINATE ABOUT THE SIMULATED GROUP DURING ITS 2 WEEKS OF OPERATION

Classification	Purpose			
	Princeton boys		Girls	
	1st week	2nd week	1st week	2nd week
	(N = 48)	(N = 5)	(N = 50)	(N = 35)
Suspicious	56.2	80.0	34.0	45.7
Unsuspicious	41.7	20.0	62.0	54.3
Indeterminate	2.1	0.0	4.0	0.0

Classification	Method			
	1st week	2nd week	1st week	2nd week
	(N = 48)	(N = 5)	(N = 50)	(N = 35)
	(N = 48)	(N = 5)	(N = 50)	(N = 35)
Suspicious	37.5	20.0	8.0	11.4
Unsuspicious	60.4	80.0	84.0	88.6
Indeterminate	2.1	0.0	8.0	0.0

in their suspicion about the purpose (exact test probability = .51 for boys, $\chi^2 = .52$ for girls) or method (exact test probability = .60 for boys and .53 for girls) of the simulated group.

Intercorrelations of Suspicion Variables

Tetrachoric correlations between the ratings were computed separately for each sex. These correlations appear in Table 3. The suspicion variables were highly interrelated. The correlations ranged from .46 to .83 for the boys and .35 to .80 for the girls. No marked tendency was apparent for higher correlations for one sex than the other, or for higher correlations between the suspicion variables derived from the same open-ended questionnaire than between corresponding variables on the two questionnaires.

Correlates of Suspicion Variables

Biserial correlations were computed, separately for each sex, between the suspicion variables and the personality scales, the intelligence test, and the conformity scores on the same procedures about which suspicion was assessed. These correlations appear in Tables 4 and 5. The patterns of significant ($p < .05$, two-tailed) correlations showed some consistencies.

Purpose and method suspicion about the simulated group had similar correlates. Both

TABLE 3

TETRACHORIC CORRELATIONS BETWEEN SUSPICION MEASURES

Measure	1	2	3	4
1. Purpose, simulated group	—	.80** (85)	.73** (75)	.46* (75)
2. Method, simulated group	.65** (80)	—	.83** (76)	.72** (76)
3. Purpose, questionnaires	.48** (73)	.35 (71)	—	.73** (76)
4. Method, questionnaires	.67** (72)	.67* (71)	.80** (74)	—

Note.—The number of subjects appears in parentheses; they vary due to missing data and to subjects classified as indeterminate. Correlations for males appear above the diagonal, those for females below it.

* $p < .05$ level.

** $p < .01$.

were, in general, negatively related to acquiescence on attitude items. Purpose suspicion correlated $-.46$ (for boys) with the Attitude Acquiescence scale, $-.33$ (for boys) and $-.30$ (for girls) with the AF Acquiescence scale, and $-.29$ (for girls) with the PF Acquiescence scale. Method suspicion correlated $-.28$ (for boys) with the Attitude Acquies-

TABLE 4

BISERIAL CORRELATIONS BETWEEN SIMULATED-GROUP SUSPICION MEASURES AND PERSONALITY SCALES, INTELLIGENCE TEST, AND CONFORMITY SCORES

Variable	Purpose		Method	
	Boys	Girls	Boys	Girls
	(N = 86)	(N = 81)	(N = 87)	(N = 79)
Fulkerson's	-.22	-.07	-.19	-.07
Acquiescence				
PRI Acquiescence	-.04	.00	-.05	-.15
Social Acquiescence	-.22	-.14	-.24	-.30
ARS	-.18	-.10	-.33*	-.26
PF Acquiescence	-.19	-.29*	-.13	-.59**
AF Acquiescence	-.33*	-.30*	-.27*	-.30
Attitude Acquiescence	-.46**	-.14	-.28*	-.36
Information-True	-.12	.03	-.22	-.20
Edwards' SD	.02	.04	.24	.15
Stricker's SD	.14	.02	.11	-.06
PRI SD	.04	.03	.14	-.10
Wiggins' SD	.06	.06	.23	.05
Marlowe-Crowne SD	.01	.00	.06	.00
Attitude SD	.29*	-.01	.25	.08
Ascendence	.20	.04	.27*	.36
Self-Esteem	.02	.08	.04	-.27
Extroversion	-.17	-.17	.10	-.19
Neuroticism	.14	-.09	-.10	.41*
Independence-Yielding	.18	.26	.24	.22
PF Authoritarianism	.08	.11	-.17	-.04
AF Authoritarianism	-.17	-.09	.12	.36
Hennon-Nelson Tests of Mental Ability	.04	.30*		
Conformity—clicks	-.44**	-.24	-.50**	-.75**
Conformity—attitude items	-.52**	-.21	-.59**	-.70**

Note.—The number of subjects of each sex varies because of subjects classified as indeterminate.

* $p < .05$.

** $p < .01$.

cence scale, $-.27$ (for boys) with the AF Acquiescence scale, $-.59$ (for girls) with the PF Acquiescence scale, and $-.33$ (for boys) with the ARS.

The correlations of these suspicion measures with the other personality scales and the intelligence test were generally not significant. (Some important exceptions were the .30 correlation between purpose suspicion and the intelligence test for girls, the .27 correlation between method suspicion and the Self-Esteem scale for boys, and the .41 correlation between method suspicion and the Independence-Yielding scale for girls.)

Both suspicion measures were, in general, negatively related to the conformity scores. Method suspicion had consistently higher correlations. Purpose suspicion correlated only with the conformity scores for boys— $-.44$ with click conformity and $-.52$ with attitude conformity; method suspicion correlated $-.50$ (for boys) and $-.75$ (for girls) with click conformity, and $-.59$ (for boys)

and $-.70$ (for girls) with attitude conformity.

Purpose and method suspicion about the questionnaires differed in their personality correlates, and there were sex differences in these correlates. The purpose suspicion of boys was positively related to SD response style—it was significantly correlated with five of the six SD scales: PRI SD scale ($r = .46$), Marlowe-Crowne SD scale ($r = .43$), Attitude SD scale ($r = .42$), Stricker's SD scale ($r = .38$), and Edwards' SD scale ($r = .36$). Other important correlates of purpose suspicion for boys were the Attitude Acquiescence scale ($r = -.32$), the Ascendancy scale ($r = .40$), the Self-Esteem scale ($r = .42$), and the intelligence test ($r = .40$).

The girls' purpose suspicion was negatively related to acquiescence on attitude items, just as their purpose and method suspicion about the simulated group had been. It correlated $-.37$ with the Attitude Acquiescence scale, $-.32$ with the PF Acquiescence scale, and $-.40$ with the AF Acquiescence scale.

Method suspicion had few correlates for either boys or girls. Among the boys, it correlated .29 with the Wiggins SD scale, .34 with the Attitude SD scale, and .29 with the Ascendancy scale; among the girls, it correlated .49 with the Independence-Yielding scale.

In sharp contrast to a sparse lack of relationship for purpose suspicion, method suspicion was consistently and negatively related to both conformity scores for both sexes. Purpose suspicion correlated only with attitude conformity among boys ($r = -.55$). Method suspicion correlated $-.53$ (for boys) and $-.36$ (for girls) with conformity in estimating probabilities, and $-.71$ (for boys) and $-.41$ (for girls) with attitude conformity.

DISCUSSION

Extent of Suspicion

The most striking findings of the present study were the wide extent and generality of suspicion. These high rates of suspicion are particularly surprising when it is recalled that the subjects were of high school age, hence representing a group presumed to typify the naive subject. It is difficult to determine

TABLE 5

BISERIAL CORRELATIONS BETWEEN QUESTIONNAIRE SUSPICION MEASURES AND PERSONALITY SCALES, INTELLIGENCE TEST, AND CONFORMITY SCORES

Variable	Purpose		Method	
	Boys	Girls	Boys	Girls
	(N = 78)	(N = 73)	(N = 78)	(N = 72)
Fulkerson's Acquiescence	-.16	-.04	.03	.07
PRI Acquiescence	.08	-.03	.09	.13
Social Acquiescence	.01	-.25	-.07	-.29
ARS	-.12	-.11	-.13	-.06
PF Acquiescence	-.09	-.32*	-.15	-.06
AF Acquiescence	-.16	-.40**	-.12	-.25
Attitude Acquiescence	-.32*	-.37*	.00	-.31
Information-True	-.08	-.06	.03	-.26
Edwards' SD	.36**	.03	.17	.08
Stricker's SD	.36**	.09	.27	.20
PRI SD	.46**	-.07	.17	.14
Wiggins' SD	.22	-.18	.29*	-.09
Marlowe-Crowne SD	.43**	-.09	.21	.04
Attitude SD	.40**	.03	.34*	.07
Ascendancy	.40**	-.02	.29*	.29
Self-Esteem	.42**	.13	.26	.32
Extroversion	-.03	-.04	.06	-.19
Neuroticism	-.09	-.03	-.04	.08
Independence-Yielding	.16	.26	.00	.49**
PF Authoritarianism	.10	-.06	-.01	-.03
AF Authoritarianism	-.27	.02	-.18	.10
Henmon-Nelson Tests of Mental Ability	.40**	.10	.27	.22
Conformity—estimating probabilities	.05	.07	-.53**	-.36*
Conformity—attitude items	-.55**	-.21	-.71**	-.41*

Note.—The number of subjects of each sex varies because of subjects classified as indeterminate.

* $p < .05$.

** $p < .01$.

whether the extent of suspicion observed in this study is atypical. Comparatively few studies report such data and, although the vast research on conformity has been extensively reviewed (Allen, 1965; Bass, 1961; Blake & Mouton, 1961; Campbell, 1961), the reviews have not been concerned with this issue. However, some bench marks can be obtained by examining several classic studies, each representing a somewhat different approach to the investigation of conformity. Sherif (1935, 1937) did not discuss suspicion in his early investigations of group influences on subjects' perceptions of an autokinetic stimulus, but in his 1937 study, which employed confederates who made predetermined responses, he reported seven subjects' introspections. None of their responses indicated any suspicion that the study was intended to discover the effects of the other person's responses on their own responses, but one subject was suspicious of the confederate: "But I sincerely believe that my partner was exaggerating the distance when he made his estimate [p. 96]."

Asch (1956), in his study of 123 subjects who judged the length of lines after confederates had made erroneous judgments, reported that "... instances of suspicion were rather infrequent. . . . [p. 29]," although

... it crossed the minds of many subjects . . . that the majority might be deliberately misleading, or that the group was following the first member who, for some unknown reason, was in error. However, this belief had the form of a fleeting hypothesis which, like many others, came and went without altering the course of action [p. 29].

Crutchfield (1955), after employing a mechanized analogue of Asch's situation, wrote that:

Of the total of 159 persons already tested in the apparatus, and questioned immediately afterwards, only a small handful expressed doubt of the genuineness of the situation. Of these not more than two or three really seem to have developed this suspicion while in the actual situation [p. 196].

Moore (1921), using questionnaires which were readministered with reports of the answers by the majority of subjects and by experts, asserted that "... the influence of suspicion was certainly not great [pp. 17-18]."

Other guidelines come from a review of recent deception studies (Stricker, 1966). Four studies predominantly concerned with conformity reported data about the extent of suspicion. The percentages of suspicious subjects were as follows: 2% of 79 subjects (Vaughan, 1964), 6% of 120 (Linde & Patterson, 1964), 18% of 60 subjects (Jones & Jones, 1964), and a "majority" of 36 subjects (Willis & Hollander, 1964).⁷

Comparisons with these studies are hazardous because of important differences in such issues as the scope of the definition of suspicion, how this information was obtained and the sophistication of the subjects. Even after taking these differences into account, it is likely that the rates in the present study were higher. One cause of the higher rates may be differences between this study and the others in their experimental designs. Although the conformity procedures in this study were modeled after commonly used ones, several distinct procedures, each involving a different deception, were employed; the other studies typically used only one conformity procedure. Exposing subjects to a variety of conformity procedures may have a cumulative effect that makes them aware of the deceptions involved. Cues from each procedure may be insufficient by themselves to arouse suspicion, but consistencies may be seen in a pattern of cues, which then triggers suspicion.

Another possible explanation lies in the enormous and ever-increasing popularity of research on conformity.⁸ This popularity has resulted in extensive reports and discussions of these studies not only in technical publications, but in elementary textbooks, widely circulated magazines, and other communication media as well. Consequently, at present,

⁷ The articles by Linde and Patterson and by Vaughan report the total number of subjects eliminated for all reasons, including suspicion. The rates cited in the text of the present article are for suspicious subjects only, and were obtained from personal communications with T. F. Linde (January 12, 1966) and G. M. Vaughan (January 7, 1966).

⁸ The citations in four reviews of the conformity literature were tabulated by year of publication. The number published before 1950, between 1950 and 1954, and between 1955 and 1959 were as follows: 6, 31, and 46 (Allen, 1965); 32, 31, and 68 (Bass, 1961); 16, 39, and 78 (Blake & Mouton, 1961); and 21, 36, and 68 (Campbell, 1961).

many potential subjects for conformity studies may already have some awareness of the nature of these studies. If this explanation is correct, it suggests that it may become increasingly difficult to undertake the usual sorts of more or less stereotyped conformity studies, particularly in literate populations. Unless variants of these studies can be found that are unrecognizable to subjects, research in this area may either have to be sharply curtailed or reinterpreted as something other than naive conformity. Of course, the potentially deleterious effects of widely broadcast information about psychological research may not be limited to investigations of conformity or other kinds of deception studies. These effects may alter the expectations and behavior of presumably "naive" subjects in a wide variety of experimental situations.

Suspicion and Conformity

Another important finding, one that was crucial in this study, was that subjects' suspicions were related to their responses to the conformity procedures—suspicious subjects conforming less. While this correlation may be a substantive one—suspicious people may be less conforming—it is more apt to be explainable in terms of a direct link between suspicion and response to conformity pressure: subjects' suspicions make them resistant to the pressures in the conformity procedure that cause others to conform. While even their rather general suspicion about the purpose of the procedure substantially affected their responses to it, suspicion about the mechanics of the deception exerted an even more powerful influence. The latter finding is not surprising, for if a person is to be maximally effective in resisting an influence, he needs to know the form that the influence will take.

Assessing Suspicion

Given the extent of suspicion and its high relation with responses to the conformity procedures, it behooves investigators in conformity studies, and perhaps other deception studies as well, to consider and assess the role of suspicion. For, in the absence of information to the contrary about subjects' suspicion, it is likely that intended measures of conformity and other dependent variables also reflect

suspicion. If suspicious subjects are detected, the best course would be to eliminate them from the study, even if they complete only a small fraction of the sample, or to analyze their data separately if that cost is sufficient. The latter procedure would have the added advantage of aiding in clarifying the nature of suspicion.

A priori considerations would suggest that the detection of suspicion is a formidable task because of several inherent measurement problems. One is that subjects' willingness to report suspicion may depend upon their perceptions of the experimenter's intentions—if they believe that he intended they be aware of the deceptions, they will report suspicion; if they believe he intended they be unaware, they will not report it (Orne, 1962). The second situation is probably the most common, and, if it did occur to this study, suggests that the rates of reported suspicion may underestimate the extent of actual suspicion. Another measurement problem is that the inadvertent use of leading questions may cause a subject to suspect deception for the first time, eliciting a report of suspicion that did not exist during the actual experiment, as well as affecting his responses to subsequent procedures employing deception. Still, the relative success in the present study with broad classifications of open-ended responses suggests that suspicion may be so potent a phenomenon that the effects of such measurement problems are subtle in comparison.

Meaning of Suspicion

Some data provide clues to the nature of suspicion, at least as it is measured in this study. Suspicion in this study was narrowly defined in terms of particulars of the experimental design. A subject was only classified as suspicious if he thought the study was concerned with changes in his responses or if he detected the mechanics of the deceptions in the conformity procedures. Thus, the measurement procedures sharply distinguished between each subject's own intrinsic suspicion about the conformity procedures, on the one hand, and such more diffuse variables as (a) a preexisting suspicion of all situations, in general, or of psychological experiments, in particular, or (b) a response tendency to re-

port suspicion, regardless of the subject's state of suspicion. While these latter tendencies were disregarded for the purposes of assessing the particularized suspicion of our experimental procedures, they might be of interest in their own right in appraising subjects' reactions in an experiment. A subject's generalized suspicions probably predispose him to seek out evidence of deception, and increase the likelihood of a veridical identification of the mechanics of the deception. In addition, even in the absence of an accurate identification of a particular deception, a generalized suspicious attitude may markedly affect a subject's willingness to cooperate by entering into an appropriate experimental set. These different forms of suspicion merit investigation as a function of both personality and situational variables.

Suspicion can, of course, be enhanced by communication among subjects. There is no evidence in this study that subjects pooled their clues and communicated their suspicions to other subjects. If such a rumor-transmission phenomenon had occurred, more suspicion would be reported by subjects during the second week than in the first week of the simulated-group procedure. In fact, the percentage of suspicious subjects during the 2 weeks did not differ significantly. This kind of transmission could have occurred, however, subsequent to the simulated-group procedure but prior to the administration of the second version of the questionnaire procedures, but this possibility cannot be appraised with the present data.

The most general findings about the personality correlates of the suspicion measures concern their negative correlations with acquiescence to scales whose content consists of social attitudes, aphorisms, and generalities, and the greater suspicion for boys than girls. These correlates may reflect passivity or naïveté in the unquestioning acceptance of the ostensible purpose of the conformity procedures, or ingratiation in concealing from the experimenter unfavorable things about the experiment (Orne, 1962). The passivity or naïveté interpretation is supported by the findings that acquiescence to scales containing aphorisms and generalities is negatively correlated with such variables as verbal abil-

ity (e.g., Bass, 1956; Messick & Fredericks, 1958) and education (Bass, 1956).

The marked relation between boys' suspicion about the purpose of the questionnaire procedures and SD response style is puzzling because of its specificity—no such relation occurred for girls, and only two of the measures of this response style were significantly correlated with boys' suspicion about the method of the same procedures. This specificity suggests that these reports of suspicion about the purpose of the questionnaire procedures had a different meaning for the boys. One conjecture is that the boys' responses may have been affected by their own previous suspicions about the simulated-group procedure. Thus, their earlier experiences coupled with exposure to the additional conformity procedures, might have cumulated sufficiently to the point where they were willing to report suspicion about the purpose of the questionnaires. Suspicions about purposes, since they require little analysis and little concrete evidence, are most susceptible to such influences. Correlations with SD response style could then arise if reporting suspicion were seen as a socially desirable or socially approved behavior, or, alternatively, if suspicion leads subjects to present themselves in a favorable light, perhaps as a means of dealing with their fear that they are being personally evaluated in the experimental situation.

REFERENCES

- ALLEN, V. L. Situational factors in conformity. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 2. New York: Academic Press, 1965. Pp. 133-175.
- ASCH, S. E. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs*, 1956, 70(9, Whole No. 416).
- BARRON, F. Some personality correlates of independence of judgment. *Journal of Personality*, 1953, 21, 287-297.
- BASS, B. M. Development and evaluation of a scale for measuring social acquiescence. *Journal of Abnormal and Social Psychology*, 1956, 53, 296-299.
- BASS, B. M. Conformity, deviation, and a general theory of interpersonal behavior. In I. A. Berg & B. M. Bass (Eds.), *Conformity and deviation*. New York: Harper, 1961. Pp. 38-100.
- BLAKE, R. R., & MOUTON, J. S. Conformity, resistance, and conversion. In I. A. Berg & B. M. Bass

- (Eds.), *Conformity and deviation*. New York: Harper, 1961. Pp. 1-37.
- BAIRD, J. W. An experimental study of decision types and their mental correlates. *Psychological Monographs*, 1914, 17(1, Whole No. 73).
- BAIM, O. G., JR. Attitude content-intensity and response. *Psychological American Sociological Review*, 1955, 20, 68-76.
- BROWN, W. Individual and sex differences in suggestibility. *University of California Publications in Psychology*, 1916, 2, 291-430.
- CAMPBELL, D. T. Conformity in psychology's theories of acquired behavioral dispositions. In I. A. Berg & B. M. Bass (Eds.), *Conformity and deviation*. New York: Harper, 1961. Pp. 101-142.
- CLAYTON, M. B., & JACKSON, D. N. Equivalence range, acquiescence, and overgeneralization. *Educational and Psychological Measurement*, 1961, 21, 371-382.
- COLLEGE ENTRANCE EXAMINATION BOARD. *Bulletin for students—1965 Preliminary Scholastic Aptitude Test*. New York: Author, 1965. (a)
- COLLEGE ENTRANCE EXAMINATION BOARD. *College Board score reports—a guide for counselors and admissions officers*. New York: Author, 1965. (b)
- COUCH, A., & KENISTON, K. Yeasayers and naysayers: Agreeing response set as a personality variable. *Journal of Abnormal and Social Psychology*, 1960, 60, 151-174.
- CROWNE, D. P., & MARLOWE, D. A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 1960, 24, 349-354.
- CRUTCHFIELD, R. S. Conformity and character. *American Psychologist*, 1955, 10, 191-198.
- EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden Press, 1957.
- FULKERSON, S. C. An acquiescence key for the MMPI. *USAF School of Aviation Medicine Report*, 1958, No. 58-71.
- GAGE, N. L., LEAVITT, G. S., & STONE, G. C. The psychological meaning of acquiescence set for authoritarianism. *Journal of Abnormal and Social Psychology*, 1957, 55, 98-103.
- GUILFORD, J. P. *An Inventory of Factors STD/C*. Beverly Hills, Calif.: Sheridan Supply Company, 1940.
- GUILFORD, J. P., & MARTIN, H. G. *The Guilford-Martin Inventory of Factors GAMIN*. Beverly Hills, Calif.: Sheridan Supply Company, 1943.
- HASTORF, A. H., & PIPER, G. W. A note on the effect of explicit instructions on prestige suggestion. *Journal of Social Psychology*, 1951, 33, 289-293.
- JACKSON, D. N. Desirability judgments as a method of personality assessment. *Educational and Psychological Measurement*, 1964, 24, 223-238.
- JONES, R. G., & JONES, E. E. Optimum conformity as an ingratiation tactic. *Journal of Personality*, 1964, 32, 436-453.
- LAMER, T. A., & NEUMAN, M. J. *Roemer's manual, the Henmon Nelson Tests of Mental Ability*. Boston: Houghton Mifflin, 1957.
- LEVIN, T. P., & FORD, C. H. Influence of the response stability on conformity behavior. *Journal of Experimental and Social Psychology*, 1966, 66, 113-119.
- MARSH, S. Response style and attitude measures from personality inventories. *Individuals and Psychological Measurement*, 1965, 21, 41-51.
- MARSH, S., & HARRISON, N. A note on the use of "acquiescence" in personality research. *Psychological Reports*, 1965, 4, 10-11.
- MILL, H. I. The social desirability influence of response set and expectancies. *American Journal of Psychology*, 1933, 32, 13-25.
- M. N. S. H. *Handbook der Psychologie*. Leipzig: Barth, 1944.
- OLMSTAD, J. A., & BLAKE, R. R. The use of administered groups to produce motivation in social norms. *Journal of Personality*, 1955, 23, 111-121.
- OLMSTAD, M. T. On the social psychology of the psychological experiment. With particular reference to demand characteristics and their implications. *American Psychologist*, 1960, 17, 775-780.
- SHIELD, M. A study of the social desirability of perception. *Archives of Psychology*, 1940, No. 187.
- SHIELD, M. An experimental approach to the study of attitudes. *Symposium*, 1957, 1, 20-28.
- STURGEON, L. J. A measure of social desirability response styles from the MMPI. *Psychological Reports*, 1966, 12(1), 419-431.
- STURGEON, L. J. *The true teacher*. Rev. Ed. 66-67. Princeton, N. J.: Educational Testing Service, 1966.
- TERMAN, L. M. A preliminary study of the psychology and pedagogy of leadership. *Pedagogical Seminary*, 1904, 11, 411-451.
- VAUGHAN, G. M. The trans-situational aspect of conforming behavior. *Journal of Personality*, 1964, 32, 633-654.
- WALLACH, M. A., & KOSAN, N. Sex differences and judgment processes. *Journal of Personality*, 1966, 27, 555-564.
- WIGGINS, J. S. Interrelationships among MMPI measures of dissimulation under standard and social desirability instructions. *Journal of Consulting Psychology*, 1959, 23, 419-427.
- WILLIS, R. H., & HOLLANDER, E. P. An experimental study of three response modes in social influence situations. *Journal of Abnormal and Social Psychology*, 1964, 69, 150-156.

(Received June 14, 1966)

SOME EFFECTS OF UPWARD MOBILE STATUS IN ESTABLISHED AND AD HOC GROUPS

DOROTHY GARTNER¹ AND MARVIN A. IVERSON

Adelphi University

Subordinate members of 4-person groups were told that 1 individual would be selected for a superior leadership position in another group, promotion depending upon meritorious performance. This orientation toward upward mobility in status was detrimental to followers' task performance in groups with prior experience in working together in experimental sessions (established groups). Among group members meeting for the 1st time (ad hoc groups), the prospect of promotion to higher status detracted from individuals' morale rather than their productivity. Comparisons were obtained with groups in which members' statuses were constant. Results are interpreted in terms of a concept of status-enhancing behavior with special reference to different types of ingroup relationships.

Empirical approaches to the problem of the effect of status striving on morale and achievement have customarily involved either the study of promotion-type incentives in field settings (reviewed by Stogdill, 1959) or the process analysis of group development (see Bradford, Gibb, & Benne, 1964). Controlled experimental analysis, especially with reference to group structure, is relatively lacking and was the focus of the present study. The particular objective was to show the effect of competition for promotion to higher status among peers who have previously interacted with each other as a group and as such have jointly established a history of successful achievements. These groups are hereafter referred to as "established" groups. It was anticipated that a policy change which would make members of established groups conscious of their individual statuses and also present them with opportunity for status enhancement would be detrimental to group productivity but would have no short-term effect on positive feeling toward the group.

A similar policy change in ad hoc groups, those which lacked a tradition of working together, was expected to detract from enjoyment of the group experience, but not necessarily affect performance in solving group problems. In order to validate these hypotheses, other groups, established and ad hoc, were compared under conditions in which

statuses were relatively constant (i.e., conditions which were intended to minimize or eliminate expectancy for promotion to higher positions). These nonmobile conditions (expectancywise) were predicted to yield higher levels of either group performance or interpersonal attraction, depending on the specific comparison.

The hypotheses were derived from a consideration of factors which generally differentiate established from ad hoc groups. Those which appeared to have special relevance in predicting reactions to a new promotion policy (one calling for the realignment of members' statuses) were differences in group-centeredness, freedom for individual action, and the degree of role differentiation.² It is known that as individuals continue to interact in becoming an established group they characteristically develop attitudes of cooperative interdependence (Jackson & Salzstein,

² Predictions were intended to apply to small work groups. It was recognized that different considerations would possibly be required in the case of other types of groups and policy changes. For example, groups which have been specifically organized to foster change or trained for survival in the face of changing environmental demands would have acquired special adaptive characteristics in responding to new policy. Further exceptions would entail proposals for change which members collectively perceive as instrumental in increasing their common group benefits (see Cartwright, 1951). Promotion based specifically on ability to foster positive feeling among group members as opposed to task achievement would also yield relationships different from those posited.

¹ Now at the State University of New York Downstate Medical Center.

1958; Lanzetta & Roby, 1957; Shaw & Rothschild, 1956). These attitudes of cooperation tend, in turn, to lead to a differentiated role structure and, also, to generate prohibitions against self-centered individualistic behavior (Shaw, 1959). In terms of this formulation, successful established groups in comparison with newly formed ad hoc groups would, then, be more highly differentiated with respect to members' roles and have more clearly defined limits on individuals' self-expressions. Members would be more group-centered in orientation. Because of these characteristics, established groups would ordinarily be expected to perform given tasks at a higher level and show more enjoyment in their experience than ad hoc groups.

These same factors which contribute to collective identification in groups, however, are to be weighted against factors which promote individual identity among members. It is reasonable to suppose that as members continue to interact and achieve success in problem solving they acquire feelings of greater freedom of action and an enlarged sphere of communication (see papers in Bradford et al., 1964). Thus a balance is struck in established groups between the inhibitions against expression of self-centered motives (implicit in the stabilized role structure) and the permissive effect of the increase in social freedom generated by successful group achievements. This balance between social freedom and restraint on self-centered behavior would be disturbed if the strength of either factor were altered.³

In the present study, it was assumed that such alteration (and thus imbalance) in established groups would result from the introduction of a promotion policy based on individual merit. The prospect of achieving higher status would encourage individual members to exercise their social freedom in expanding their range of behavior and, at the same time,

reduce their conformity to existing role restrictions. Coordinated effort would be weakened as a consequence, and overall group performance would be lowered. Group morale would initially, however, remain unchanged. The pleasure which members would find in the self-styling of new behavior and their anticipation of being promoted would reinforce the enjoyment derived from their previous successful group interaction.

In ad hoc groups, a policy of competitive promotion in status would conceivably not detract from group performance. Such a policy could not disrupt a regularized mode of group interaction since none could exist. Moreover, the prior demands on members of newly formed groups to settle quickly on a work routine necessary for successful task completion would be expected temporarily to hold in check any tendency for self-centered behavior. Unlike established groups, members of ad hoc groups could not take coordinative effort for granted. At the same time, however, competition for promotion in status, because benefits accrue to only one person, would be expected to retard the development of positive feelings of intimacy and trust among newly affiliated group members. And, there would be no prior group cohesiveness to offset this effect.

METHOD

Overall Design

The experiment involved three preliminary sessions of an hour each designed for developing a history of successful interaction among peers for 16 groups. This treatment defined "established" groups. Members of half of the ad hoc groups (eight groups in number) went through the same preliminary procedure employed for established groups except that at each successive meeting they interacted with other individuals whom they were meeting for the first time. The remaining eight ad hoc groups met only once to undergo the final experimental procedure.

The second phase of the experiment was a final hourly session in which all groups were required to perform a common critical task. Half of the groups were given an orientation which was designed to arouse an interest in status enhancement (i.e., promotion to a position of leadership). The other half

³ This relationship has been discussed by Stogdill (1959) in terms of the balance between authority and responsibility which is essential to high group achievement. Such balance according to him is especially unstable when a new policy poses conflict or alteration of individuals' statuses. The susceptibility of organized, cohesive groups to disruptive influences has also been referred to by French (1953).

⁴ Kidd (1958) has observed, for example, that unacquainted college students when assigned a group task agree on a work procedure and move quickly toward a behavioral asymptote within a single hourly session.

received instructions which were devoid of references to status mobility. Group performance on the task and individuals' ratings of their feelings of comfort and congeniality in participating with other members were recorded for this session. These measures provided the primary data in the testing of hypotheses.

The established and ad hoc treatments were combined factorially with the presence and absence of opportunity for promotion (mobile-nonmobile status). A total of 128 undergraduate women enrolled in introductory psychology courses⁵ were convened to form four-member groups. The experimenter phoned each individual student beforehand to make certain that she was unacquainted with other members of her group and to work out a suitable meeting schedule. Eight groups (or 32 subjects) were assigned at random to one of the four combinations of type of group and status orientation. When recruited, all group members were told that they would be paid for their time and that payment would follow the last experimental session. There was no attrition from the original sample selected for investigation.

Procedure for Established Groups

The 16 groups identified as established groups were told at the outset that they would meet four times, always with the same group members. During the first three hourly sessions which were specifically designed to establish a group tradition, groups successfully completed four tasks assigned by the experimenter. Tasks were selected in terms of affording a range of problem-solving activities and requiring collective interaction for their successful completion. They were also of a type which was almost entirely self-administering for the group, thus allowing the experimenter to remain in the background as a nonparticipating observer. They are briefly described as follows:⁶

1. Twenty questions. One group member was randomly chosen as informant and given a card on which the name of a commonplace object was typed. The goal was for the other three members to guess the object by asking questions which could be answered by "yes," "no," or "sometimes." They were allowed a maximum of 20 questions, but each question had to be one on which the three members agreed. When the object was guessed correctly, as it almost always was, the role of informant was rotated randomly.

2. Student ratings. Group members were required to discuss a standard list of 12 student offenses obtained from Zeff and Iverson (1966). The goal was to derive a group consensus in rank ordering the 12 items in terms of their relative seriousness. The list took an average of 25-30 minutes to scale.

3. Vigotsky Concept Formation Test. This test of concept formation was administered as prescribed by

⁵ The authors are indebted to course instructors at Brooklyn College who assisted in the recruitment of students to serve as group participants.

⁶ For a fuller description of the tasks, see Gartner (1962).

Hanfmann and Kasanin (1937). An additional instruction was that group members necessarily agree collectively on the sorting of blocks into four categories and their solution principle. The experimenter provided standard cues in those groups which did not obtain a solution in the first 15 minutes. All groups derived an adequate solution in less than a half hour.

4. Windmill construction. The group was given a construction set and a diagram of the front and side view of a windmill. They were told to work together in building the object represented in the drawing. Since the diagram lacked certain detail, a variety of possible constructions could be considered adequate. This task required ordinarily an hour to complete.

The first session was the same for the 16 established groups. The 20-questions procedure was employed because it was familiar to most participants, and it rapidly encouraged social interaction among unacquainted persons. Tasks for the second and third sessions were varied at random. Because of time required for completion, student ratings and the Vigotsky Concept Formation Test were combined to comprise one session, but were alternated at random with the windmill-construction task to comprise another session of the group.

At the end of each tradition-building session, group members individually supplied a sociometric rating of leadership and filled out a questionnaire rating of their reactions to the group atmosphere and the tasks given them.

Ad Hoc Treatments

Of the 16 ad hoc groups, 8 were seen only during the administration of the final critical task. Members of the remaining 8 groups participated in four hourly sessions, but were rotated from group to group in such a way that they never interacted with another member for more than 1 hour, including the final meeting. With the exception that individuals were performing as members of a new group at each session, they were given the same tasks and general instructions as the established groups. Additional precaution in the rotation of members was that no individual undertook a given task more than once. This special treatment of ad hoc groups made it possible later to verify that the effects of the treatment accorded established groups depended upon meeting with the same group members rather than mere experience with a particular set of tasks.

Status Orientations

At the beginning of the final experimental session, all groups were told that their task would require a leader. In the eight ad hoc groups with no previous laboratory experience, the experimenter named the leader at random and informed the group of this fact. In the eight additional ad hoc groups and the 16 established groups, members which were meeting for the fourth time, the experimenter ex-

plained that the choice of leader for the session was based on previous sociometric ratings—which, in fact, it was.

After designating a leader, the experimenter in the sense of introducing a new policy notified the other three members in half of the established groups and half of the ad hoc groups (in turn, half of each ad hoc treatment) of their opportunity for promotion to a leadership position in other groups. They were told that other groups were meeting at the college and that someone from their group depending on her performance would be chosen to fill a vacant leadership position. Although, not stated explicitly, it was implied that the "chosen" one would have the chance to increase her earnings. The experimenter stated that a leader could serve in only one group. The latter comment was inserted to remove any implications of demotion in status and to eliminate interlevel rivalry between leader and followers within the same group. By disqualifying the present group leader, the investigators attempted to define a condition of competitiveness among the three subordinate members.

The remaining established and ad hoc groups were identified experimentally as being nonmobile in status. After designating a leader for the current session, the experimenter proceeded with the standard instructions for the critical task.

Critical Task

The critical task involved a mapping problem similar to one described by Shaw (1959). Subjects were shown a blackboard graph of 100 squares with intervals of from A to J on the abscissa and 1 through 10 on the ordinate. The experimenter had prepared ahead of time diagrams consisting of four concentric circles similar to a bull's-eye target, except that the circles were irregular, and the bull's-eye could be located anywhere on the chart. Different values were assigned to different areas such that the perimeter area had the smallest value and the center area or the bull's-eye had the highest value. A typical problem diagram is illustrated in Figure 1.

On each trial, the group without seeing the experimenter's diagram was required to decide upon a pair of coordinates (e.g., C5, D9, etc.), the object being to guess as accurately as possible the location of the bull's-eye, thereby earning maximum points for the group. The same coordinate could be chosen as many times as desired. Consequently, the fewer trials required to locate the bull's-eye, the greater the group's accuracy in mapping and the more points members could accumulate on a given problem.

Each group had to map 5 diagrams with 10 trials per problem, requiring 50 choices in all. The three subordinate group members were instructed to suggest their choice of coordinates to the group. The leader was not allowed to initiate solutions, but retained the deciding voice in rejecting or selecting among solutions offered by followers. Leaders were further instructed to do whatever seemed necessary

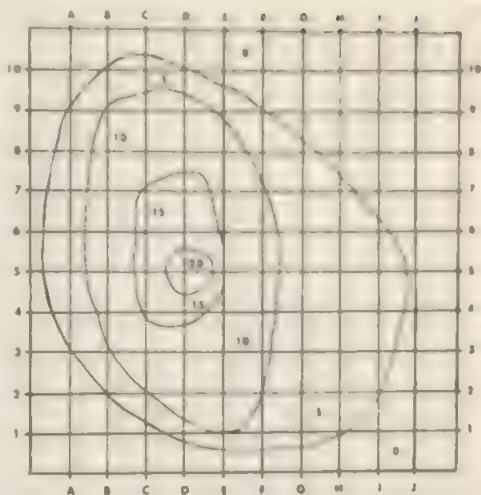


FIG. 1. Typical diagram used for the critical task.

to facilitate the maximum score in the shortest possible time. They had no foreknowledge of the correct coordinates. The leadership role was prescribed as relatively nonparticipatory in contrast to follower roles in order to equalize the effectiveness of random with sociometrically nominated leaders.

As each choice of coordinates was decided, the experimenter marked the value of the chosen area on the blackboard for all subjects to see. The experimenter also entered the scores on a worksheet and, at the end of each problem, recorded the total time taken by the group for mapping.

The final procedure was terminated by having members individually complete a questionnaire. Items were designed to determine, where applicable, whether or not subjects had perceived changes in group atmosphere over the four sessions, their degree of interest in becoming leaders, their satisfaction with the leaders in the final session, and their awareness of individually prominent behavior in themselves and other group members. In dismissing groups, the experimenter briefly answered all questions and clarified for groups with upward mobility the experimental objectives of the status orientation.

RESULTS

Performance measures and satisfaction ratings obtained from ad hoc groups which met once for the critical task were compared with those derived from ad hoc groups in which members had participated in three preliminary sessions but each time in different groups. A 2×2 analysis of variance with status mobility as a second classification factor (four groups in each cell) yielded nonsignificant F ratios (p 's $> .20$). Therefore, effects obtained with the two types of ad hoc groups were re-

garded as homogeneous, and data were pooled in further statistical analyses.

Effects of Major Variables on Performance

The major analyses of performance data were carried out according to a mixed model, established-ad hoc groups and mobile-nonmobile status comprising the independent factors, and the five problems on the critical task providing the repeated variable.

Analysis of variance of overall time measures on the mapping problems failed generally to show either reliable main effects or statistical interaction. All F ratios had p values greater than .20. With respect to accuracy scores in mapping, however, a significant interaction between type of group and mobile status was obtained ($p < .01$). This interrelationship is illustrated in Figure 2 and may be ascertained from the descriptive statistics in Table 1 and the summary of analysis of variance in Table 2. Established groups with no status orientation were the highest in achievement. The established groups with upward orientation in status were less successful and performed similarly to ad hoc groups under either mobile or nonmobile conditions. Loss in problem-solving accuracy in established groups with opportunity for individual promotion was in accord with the hypothesis.

The reliable main effects of type of group and status mobility as shown in Table 2 appeared to derive from the interaction effect and were therefore given no further consideration. Variance contributed by problems was statistically nonsignificant (p 's $> .10$).

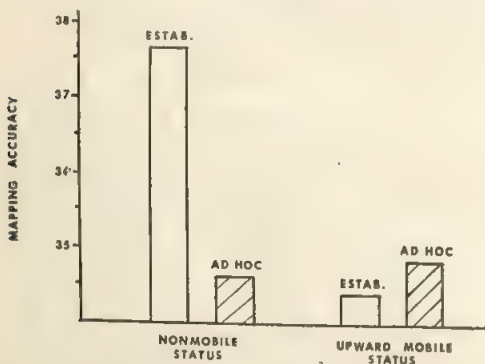


FIG. 2. Interrelated effects of type of group and status mobility on accuracy scores.

TABLE 1
MEANS AND STANDARD DEVIATIONS
OF ACCURACY SCORES

Group conditions	<i>M</i>	<i>SD</i>
Established-Upward Mobile Status	34.4	2.1
Established-Nonmobile Status	37.6	2.0
Ad Hoc-Upward Mobile Status	34.8	2.3
Ad Hoc-Nonmobile Status	34.6	2.9

Note.— $N = 8$ for all group conditions.

Effects of Major Variables on Ratings of Positive Group Feeling

Table 3 presents means and standard deviations of ratings which were derived from a self-anchored graphic scale administered at the end of the critical task. Subjects were individually required to evaluate their feelings of comfort and friendliness during the group performance. The high end of the scale represented discomfort and restraint. Members' ratings were summed for each group, and the combined value was treated as a single score. Data were assembled in terms of a 2×2 fixed model.

An analysis of variance as summarized in Table 4 revealed a significant interaction ($p < .025$) between the variables of type of group and status mobility. Established groups expressed relatively high friendliness and comfort irrespective of their status orientation. In the ad hoc groups, the prospect of individual promotion in status (as opposed to conditions where such opportunity was absent) appeared to result in less feeling of congeniality in the group (or greater restraint). This finding corresponded to hypothesis and is illustrated in Figure 3.

TABLE 2
SUMMARY OF ANALYSIS OF VARIANCE
OF ACCURACY SCORES

Source	<i>df</i>	<i>MS</i>	<i>F</i> ratio
Status mobility (B)	1	66.31	5.86*
Type of group (C)	1	91.50	8.09**
B \times C	1	120.77	10.60**
Error (b)	28	11.30	
Problems (A)	4	6.11	.28
A \times B	4	38.60	1.78
A \times C	4	28.43	1.31
A \times B \times C	4	6.15	.28
Error (w)	112	21.60	

* $p < .025$.

** $p < .01$.

TABLE 3

MEANS AND STANDARD DEVIATIONS OF RATINGS OF POSITIVE GROUP FEELING

Group conditions	M	SD
Established-Upward Mobile Status	2.7*	.73
Established-Nonmobile Status	2.8	.53
Ad Hoc-Upward Mobile Status	3.7	.43
Ad Hoc-Nonmobile Status	2.8	.50

Note.—N = 8 for all group conditions.

* Low end of the scale indicates positive group feeling.

The significant main effect associated with type of group was given no special consideration in view of the statistical interaction with status mobility.

Questionnaire Findings

Questionnaire ratings as supplied by subjects at the end of each group session were analyzed to verify the effectiveness of operational procedures. Under the conditions of upward mobile status, follower subjects were asked after completing the critical task whether or not they were interested in becoming leaders in another group. In the established-mobile groups, 75% gave definite positive answers. In ad hoc groups, 71% responded positively. These data indicated that the large majority of individuals in both established and ad hoc groups was sensitive to the experimenter's status orientation.

As a check on reactions to the leader chosen for the final session, nonleaders were asked to rate the appointee as good, mediocre, or poor. In all instances, the group leader was judged good. Consequently, leaders were regarded as equally effective, irrespective of whether they were randomly selected or sociometrically chosen.

After each of the three preliminary task sessions, group members recorded individually

TABLE 4

SUMMARY OF ANALYSIS OF VARIANCE OF RATINGS OF POSITIVE GROUP FEELING

Source	MS	F ratio
Status mobility (A)	1.30	3.61*
Type of group (B)	2.10	5.83**
A × B	2.26	6.28**
Error (b)	.36	

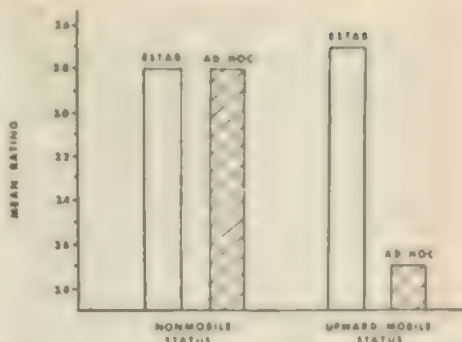
* $p < .10$, $df = 1/28$.** $p < .025$, $df = 1/28$.

FIG. 3. Interrelated effects of type of group and status mobility on ratings of positive feeling toward the group. (Low scores represent positive ratings.)

their impressions of (a) other group members, (b) the group atmosphere, (c) reaction to the task, and (d) involvement of the group in the task. With successive meetings, all groups in which membership remained the same (i.e., the established groups) tended to view the group and the tasks in an increasingly favorable light and to rate more highly the degree of cooperation and involvement in each other and in the tasks. As shown in Table 5, no consistent trend characterized ratings from the groups with changing membership (i.e., the experienced ad hoc groups). Because the order of task assignments was varied from group to group, the progressive increase in group-centeredness in established groups necessarily related to repeated meetings with the same persons rather than to the particular tasks undertaken.

Evidence of role differentiation in established groups was supplied by the fact that members in all cases by the third session independently agreed in a sociometric rating as to their group's leader.*

TABLE 5

MEAN QUESTIONNAIRE RATINGS FOLLOWING PRELIMINARY GROUP MEETINGS

	Session	Q.1 Group members	Q.2 Group atmosphere	Q.3 Task	Q.4 Group to task
Established groups N = 16	1	4.2	3.5	4.1	4.3
	2	4.5	4.3	4.3	4.7
	3	4.7	4.6	4.4	4.9
Ad hoc groups N = 8	1	4.5	3.8	4.2	4.5
	2	4.6	3.7	4.2	4.7
	3	4.3	3.9	4.2	4.6

Note.—Positive ratings are represented by higher numbers.

DISCUSSION

The results, although limited in scope, bear on the broader issues of adaptation to policy change and status-seeking behavior in newly formed and in established groups. New policy which directed individuals' interest toward enhancement of their statuses was seen to have a deleterious effect on the quality of performance in established groups but did not have a counterpart in a loss of positive group feeling. When policy merely reinforced existing statuses, the factor of a tradition of successful group achievement appeared to facilitate accuracy in carrying out another task assignment. In newly formed, traditionless groups, the immediate effects of status orientation were more evident in retarding the development of group feelings of intimacy and trust rather than interfering with performance. In a general way, the evidence indicated that promotional opportunity in status failed to encourage positive group sentiment. Indeed, by implication, the prospect of status enhancement appeared to foster an orientation toward individual, self-centered goals—ones which are incompatible with feelings of collective identity.

Status-Enhancing Behavior

Of interest in considering results of the present investigation in terms of previous experimental studies of status was a comparison of behavioral effects which are said to be status enhancing and those which are status protective. In its functional significance for the individual, a gain in status assuredly differs from a loss in status (relative to a control condition of no change), and thus responses would differ with the prospect of promotion as compared to threat of demotion. Striving for status enhancement and status protection would be expected to take different behavioral forms.⁷ For example, investigators who have explicitly examined the threat of

downward mobility have noted certain constriction in communication (Kelley, 1951) and concern with establishing a conforming, protective identity within a hierarchy (Zeff & Iverson, 1966).

On the basis of findings in the present study, behavior in a potentially status-enhancing situation, instead of being constrictive, seemed to comprise an expansive mode of task effort and one designed to facilitate movement on the part of an individual within a group. With the prospect of individual promotion—an occasion which expressly called for status-enhancing performance—there was some evidence that members of established groups would exercise more latitude in trying out new, individually prominent behavior. In the final questionnaire, a substantial number of "established-mobile" subjects associated changes in their own behavior (33% of the total sample) and the behavior of other members (44%) with the possibility of being selected for promotion. Subjects stated that they "tried to make many suggestions," or noted that one member "sat on the table to become more noticeable," or "she giggled and laughed a lot to gain attention." These descriptions of behavioral shift during the critical task session corresponded to a log of such behaviors kept by the experimenter.

In drawing attention to their individual performances, however, members necessarily weakened coordinative control and lowered the quality level of the overall group performance. They could do so and still be assured of respect from their associates, at least temporarily, because of their effective past relationships. Evidence for such respect was present in the questionnaire evaluations from earlier sessions. Members' continued enjoyment in group participation could be regarded as a carryover feeling and, also, as a consequence of their increased behavioral expression rather than as an indication of greater collective group sentiment.

In ad hoc groups wherein members had no common past experience to override feelings of initial distrust, the prospect of upward movement in status appeared to prolong their feelings of mutual restraint and aloofness but in such a way that it did not reduce group productivity. The relative absence of positive

⁷ Status protection and status enhancement may be seen as socially determined forms of ego defense and ego enhancement. Evidence regarding the constrictive effects of ego defensiveness on behavior and the facilitative effects of ego enhancement (see Iverson & Reuder, 1956) would then be applicable in predicting individuals' responses to prospective change in status.

group feeling was presumably balanced by increased task effort, but in contrast with established groups such effort was attuned to the expanded effort of other group members. Such groups were, thus, enabled to perform at the same level as ad hoc groups in which statuses were nonmobile.

The latter observations were consistent with findings of other experimental studies of ad hoc groups. In analyses of communication content, Kelley (1951) and later Cohen (1958) reported data indicating that group members with opportunity for competitive promotion showed relative disinterest in developing close ingroup ties, but also had heightened sensitivity to issues which bore on their present jobs. In a similar sense, Zeff and Iverson (1966) noted that individuals with potential upward movement tended to focus on the group's task but to make relatively little use of their own peer group as a basis for identifying themselves statuswise.

REFERENCES

- BRADFORD, L. P., GIBB, J. R., & BENNE, K. D. (Eds.) *T-group theory and laboratory method*. New York: Wiley, 1964.
- CARTWRIGHT, D. Achieving changes in people: Some applications of group dynamics theory. *Human Relations*, 1951, 4, 381-392.
- COHEN, A. R. Upward communication in experimentally created hierarchies. *Human Relations*, 1958, 11, 41-54.
- FRENCH, J. R. P. The disruption and cohesion of groups. In D. Cartwright & A. Zander (Eds.), *Group dynamics*. Evanston, Ill.: Row, Peterson, 1955. Pp. 121-134.
- GUINER, D. *Leadership, group status, and group performance*. Unpublished doctoral dissertation, Adelphi University, 1966.
- HANFMAN, E., & KASANIN, J. A method for the study of concept formation. *Journal of Psychology*, 1937, 3, 521-540.
- IVERSON, M. A., & REUDER, M. E. Ego involvement as an experimental variable. *Psychological Reports*, 1956, 2, 147-181.
- JACKSON, J., & SALTZSTEIN, H. The effect of person-group relationships on conformity process. *Journal of Abnormal and Social Psychology*, 1958, 57, 17-24.
- KELLEY, H. H. Communication in experimentally created hierarchies. *Human Relations*, 1951, 4, 39-56.
- KIDD, J. S. Social influence phenomena in task-oriented group situations. *Journal of Abnormal and Social Psychology*, 1958, 56, 13-17.
- LANZETTA, J., & ROBY, T. Group learning and communication as a function of task and structure demands. *Journal of Abnormal and Social Psychology*, 1957, 55, 121-131.
- SHAW, M. E. Some effects of individually prominent behavior upon group effectiveness and member satisfaction. *Journal of Abnormal and Social Psychology*, 1959, 59, 382-386.
- SHAW, M. E., & ROTHENBILD, G. H. Some effects of prolonged experience in communication nets. *Journal of Applied Psychology*, 1956, 40, 281-286.
- STOGDILL, R. M. *Individual behavior and group achievement*. New York: Oxford University Press, 1959.
- ZEFF, L. H., & IVERSON, M. A. Opinion conformity in groups under status threat. *Journal of Personality and Social Psychology*, 1966, 3, 383-389.

(Received December 14, 1965)

AN EXAMINATION OF THE RELATIONSHIP BETWEEN ORDINAL POSITION, PERSONALITY, AND CONFORMITY:

AN EXTENSION, REPLICATION, AND PARTIAL VERIFICATION:

EDWARD E. SAMPSON

AND

FRANCENA T. HANCOCK

University of California, Berkeley

Crystal Springs School, Hillsborough, California

Data obtained from high school children of 2-sibling families representing all possible combinations of ordinal position, S sex, and sibling sex indicate that the variable of ordinal position accounts for significant differences in *n* Achievement and *n* Autonomy. Firstborns have higher *n* Achievement and *n* Autonomy, tend towards lower test anxiety, and, if male, manifest more achievement-related conformity than second borns.

In 1962 Sampson presented data from three separate studies which led him to conclude that: (a) Firstborns have a higher need for achievement than later borns; (b) firstborn females are *more* resistant to influence than later-born females; (c) firstborn males are *less* resistant to influence than later-born males (Sampson, 1962). Those conclusions were based on separate studies using separate populations of subjects in rather different situations of social influence. The inferences which related birth order through personality to conformity, therefore, were sufficiently tenuous to lead the author to state that "... a strict test of that hypothesis would involve placing males and females into the *same* influence situation and measuring their resistance to influence [p. 159]." In large part this provides the focus of the present article.

Since 1962, two extensive reviews of the literature have appeared (Sampson, 1965; Warren, 1966), and several new and enlightening research articles have emerged (cf. Becker, Lerner, & Carroll, 1966; Carrigan & Julian, 1966). Both review articles reported several consistencies in findings relating ordinal position to personality variables and to certain outcome variables, including conformity; these consistencies made it apparent that the renaissance in this area of investigation was not to be understood so simply as just

another unfruitful research fad. In his review, however, Sampson urged future researchers to attend to several neglected methodological and psychological matters by (a) selecting subjects with some care to their birth order, family size, sibling sex, and sibling-age spacing, (b) separating in analysis firstborns from only children, (c) avoiding the psychologically vague operation of lumping everyone else together into the catchall category of later born. In addition, it was suggested that this methodological concern over the sample be supplemented by an effort to measure the presumed intervening psychological and social processes that are always posited to account for ordinal position differences.

Although few, if any, of the recent ordinal position studies have taken careful note of most of the methodological points, several have provided additional valuable insights into the complex relationship between birth order and conformity, the topic of this present paper. The surprise in the data of the 1962 study was the finding that firstborn females conformed *less* than later-born females, an apparent exception to the general statement that firstborns are more conforming than later borns. The relationship between ordinal position and conformity and this one apparent exception appears to be based in part upon the simple working assumption that conformity is a rather undifferentiated variable. The studies of Becker et al. (1966) and of Carrigan and Julian (1966) lead one to question this assumption. Basing their work on the distinction proposed by Deutsch and

¹ This study was supported by Research Grant M-5747-A from the National Institutes of Health, United States Public Health Service. Preliminary results were mentioned in Sampson (1965). This present report is one part of a larger study which was the doctoral dissertation of Hancock (1965).

Gerard (1955), between normative and informational influence, Becker et al. concluded that whereas firstborns are more responsive to normative influences, later borns are more affected by informational influences. This finding is consistent with the theoretical interpretation which suggests that firstborns are more affiliatively dependent, need others for social support, and thus should be susceptible to a normative type of influence situation. Carrigan and Julian presented data which not only confirm the importance of differentiating between kinds of conformity situations, but which also indicate an interaction with the variable of sex that is somewhat similar to that reported by Sampson in 1962. Their data indicate that when affiliative anxiety is aroused through an experimental manipulation, firstborn males conform more than later-born males, but there is no significant difference in conformity behavior for the females.

Although the preceding studies leave many ordinal position variables uncontrolled and unexamined, they provide a direction which cannot be disregarded in attempting to understand the connection between birth order and so complex and differentiated a variable as conformity. The nature of the influence situation, the kinds of needs it arouses, and the sex (and probably age and social class as well) of the subjects provide several complex interactions which must temper one's conclusion that firstborns conform more than later borns. In the study that follows, the authors made an effort to explore systematically the nature of this relationship by differentiating and controlling the variables of ordinal position, by studying their relationship with several variables of personality, and by relating them to a defined type of conformity situation in which all subjects participate. Thus, in this study we are able to provide a substantial replication of the earlier study (Sampson, 1962) while at the same time examining in a more controlled setting some of the intervening processes so tentatively inferred from that earlier work.

METHOD

In general outline, the study involved obtaining a selected sample of students at a local high school²

² We would like to thank D. K. Freudenthal, research director of the Berkeley Unified School Dis-

TABLE 1

Distribution of Subjects by Birth Order,
Subject Sex, and Sex of Sibling

Category	Cell Number	N	%
Firstborn boy with younger sister	(M,M)	24	3
Firstborn boy with younger brother	(M,F)	26	4
Second-born boy with older sister	(M,M)	24	5
Second-born boy with older brother	(M,F)	20	6
Firstborn girl with younger brother	(F,M)	32	7
Firstborn girl with younger sister	(F,F)	26	8
Second-born girl with older brother	(F,M)	26	9
Second-born girl with older sister	(F,F)	26	10

and placing them into several testing situations, including one involving the experimental manipulation of social influence. In order to minimize the effect of any work done with a variable such as ordinal position, it is interesting that in the nature of the sample used as well as the specific methods of measurement and experimental treatments employed, we shall first present several details of the sample.

Sample

In the spring of 1961, each student in the eleventh grade of the selected high school received a questionnaire in his home on a card listing the names, sex, and age of himself and his siblings. Shortly thereafter additional students from Grades 10 and 12 were provided with a similar questionnaire. Letters describing the study and inviting the student to take part were mailed to all students who were members of one- and two-child families and whose sibling was not more than 5 years separated from him in age. The initial questionnaire indicated that about 400 students met these criteria. Attendance, which was voluntary, at two 45-minute sessions in a small lecture hall at the high school just before and after school hours, was the final criterion for inclusion in the sample.

Two hundred and fifty-one students met all criteria for inclusion and thus formed the sample. Table 1 presents the cell by cell analysis of the sample by the three ordinal position variables examined, birth order, subject sex, and sibling sex.

Based on their responses to one of the questionnaires, it was possible to examine relevant characteristics of the sample and determine whether or not the 10 cells of the sample differed significantly

trict, E. J. Curtice, principal of Berkeley High School, and the many school counselors, students, and parents for their excellent cooperation throughout all phases of this project.

On the usual array of control variables; that is, age, grade in school, sibling spacing, father's occupation, and education. An examination of the relevant data indicated that between-cell variation in age, grade, and spacing was so slight as not to affect the variables under study. Father's occupation and education, employed as the measure of social class, although not as fully consistent across all cells as one would have liked, do not appear to vary so much as to render meaningless the relationships examined.

With regard to full sample averages, the authors' data indicate that the mean age of the subjects is 16.2; their mean school grade is low 11; the mean spacing between siblings is 2.7 years, with 97% of the sample being described by a spacing of no more than 5 years; and finally, based upon Warner's 7-category system (Warner, Meeker, & Eells, 1949), the average occupational class rating is 2.2.

In passing, it is of some interest to note that the sample included a slight overpopulation of firstborns. Whereas nearly equal numbers of first and second borns composed the total subject pool, of those who volunteered, a slightly greater percentage were first-borns than second borns. This is consistent with several studies that have suggested that under certain conditions firstborns are more likely to volunteer than are later borns (Capra & Dittes, 1962; Suedfeld, 1964; Weiss, Wolf, & Wiltsey, 1963).

*Instruments and Procedures*³

All subjects were mailed a general description of the project, scheduling information, and several short questionnaires which they were asked to complete and mail back before they arrived for their two experimental sessions. The description mentioned that this was a research project conducted by the Department of Psychology at the University of California under the sponsorship of the National Institutes of Health. The study was described as consisting of an investigation of attitudes and personality and of their relation to child-rearing procedures. Anonymity was assured. The questionnaires which the subjects were asked to complete at home consisted of a slight modification⁴ of the measures of achievement and affiliation from the Edwards Personal Preference Schedule (1953) and a slightly modified version of the Winterbottom (1958) scale of independence training. Subjects were to schedule themselves for two out of six possible testing times which took place in one of the small lecture halls at the high school and which were held before school in the morning and after school in the afternoon on 3 consecutive days in the middle of 1 week. Each session was about 45 minutes in length. The smallest

session had 50 subjects in attendance; the largest had 118.

Two different test booklets were administered to the subjects. These booklets contained the measures of various socialization (cf. Footnote 3) and personality variables, as well as the measure of conformity to peer influence. An experimenter plus two assistants worked each session.⁵ The experimenter was a male graduate student who introduced the project once again as a university-sponsored research study and assured the subjects of their anonymity. He conducted all sessions, gave all directions, and read aloud every item they were to answer.

Conformity. The conformity task involved subjects' estimates of the number of circles drawn on a large cardboard poster and held up before them, and their estimates of the height of the center line of an equilateral triangle, also drawn on a large cardboard poster and held up before them. The entire task was presented as a study of the effects of time on estimation abilities. The specific instructions stressed the importance of the ability to estimate well, and urged everyone to do his best.

The card containing the circles was displayed for approximately 15 seconds. The students then made their estimates. The next task—estimating the height of a line drawn from the base to the apex of a triangle—was introduced with instructions very similar to those employed for the circle-estimations task. The triangle was then displayed for approximately 15 seconds, and the subjects made their estimates.

All estimates were written on a piece of paper provided as part of the first testing booklet. The subjects placed their names on this sheet, tore it out of the booklet, and passed it in. The assistants collected these sheets and took them to the front of the lecture room at the desk behind which the experimenter was standing. While the experimenter continued with the other measures included as part of this testing session, the assistants were rather obvious as they appeared to count and supposedly average the estimates.

After a period of about 15 minutes had elapsed, during which time the subjects were completing other questionnaires, the experimenter introduced the second part of the conformity test which had the students once again view each card, now for about 5 seconds, and make their estimates. This second estimation was presumably part of the study of the influence of time on judgmental ability. Just before each reestimate was made, one of the assistants handed the experimenter a note, supposedly containing the computation of group norms on their previous judgments. The experimenter casually mentioned that most people thought that there were

³ We shall mention here only those measurements which are of central relevance to this report. A fairly extensive array of measurements was employed. The more complete report may be found in Hancock (1965).

⁴ This modification involved eliminating a few items which made reference to topics felt to be inappropriate for use in the high school.

⁵ We would like to thank Dwight Harsbarger who served as the experimenter and Virginia Peek who served ably as a general assistant and all-around helper on the project. Wayne Sailor and Sheldon Berkowitz are due many thanks for their help in the complex computer operations.

TABLE 2
ANALYSIS OF VARIANCE

	M	F	M:M	M:F	M:M	M:F	Sex	Age	Task	Item
\bar{X}	27.46	22.42	26.32	21.33	24.10	20.10	11.00	20.00	11.76	11.81
S	13	12	19	20	20	19	20	19	14	19
Analysis of variance										
Source of variance	Residual		SS		df		F		p	
Ordinal position (A)	7508.642		75.086		1		3.13		.08	
Sex (B)	7692.362		168.362		1		15.00		.001	
Sex of sibling (C)	7579.308		28.308		1		.27		.61	
A \times B	7608.401		137.401		1		11.69		.001	
A \times C	7524.754		53.754		1		.45		.50	
B \times C	7493.964		3.964		1		.04		.83	
A \times B \times C	7479.682		8.682		1		.08		.78	
Error	40.936				180		1.00			

* High number indicates more conformity.

** $p < .05$.

about 53 circles on the card or that the triangle was about 18 inches high.

These two values—53 circles and 18 inches—were not the actual average value given by the group, but rather were values which approximated those given by pretest groups.

A conformity index was computed based on the change in judgment from the prenorm to the post-norm estimate on both tasks combined. This index took into consideration both the direction of the change—whether towards the norm or away from it—as well as its amount.

Personality measures. In addition to the measures of achievement and affiliation from the Edwards Personal Preference Schedule which were previously mentioned, the measure of autonomy was also used. Two further tests were employed: a shortened 11-item version of the Mandler-Sarason Test Anxiety scale (Mandler & Cowan, 1958), using the items which refer to tests and testing situations in general, and three items from the French Test of Insight (1958) designed to measure achievement imagery.

RESULTS

The basic design of the study involves all possible combinations of three independent factors—subject's ordinal position, subject's sex, and sibling sex—and a series of dependent variables. If the two cells involving only children are excluded, it is possible to place the completed design of the remaining eight into a complex three-way analysis of variance. Additional analyses allowing a comparison of the results for the only children with the remainder of the sample were also conducted. Where it was felt necessary to examine sev-

eral cell-by-cell differences, the method of Scheffé (1959) as suggested by Hays (1963) was used. As the major dependent variable was conformity, those results will be examined first and then we will work our way backwards theoretically.

Conformity

Table 2 presents the mean conformity index for all 10 cells and the analysis of variance conducted on the eight dyadic cells.* The analysis of variance indicates the main effect of subject's sex to be significant at less than the .05 level, and the interaction between ordinal position and subject's sex to approach, but not quite reach, an acceptable level of statistical significance ($p < .10$). An examination of the means indicates that the sex difference in which males conform more than females is almost entirely attributable to the high level of conformity of the firstborn males. Among the female groups, regardless of ordinal position or sex of sibling, the level of conformity is approximately the same. Separate cell-by-cell analyses indicate that among the males, firstborns conform significantly more than second borns ($p < .01$), with the difference between M:M and M:F being highly significant ($p < .001$).

*Variations in the N s from comparison to comparison are to be taken as a function of incomplete data for a particular subject or set of subjects on a particular item.

When the conformity for the only-child groups is examined, we note that the sex difference which emerged on the analysis of variance is upheld. Only-child males fall at the one extreme of high conformity, whereas only-child females fall at the other extreme of low conformity.

It appears that these results offer some confirmation for the findings of the earlier study (Sampson, 1962) in which males who were firstborn conformed significantly more than males who were later born, whereas the direction was slightly reversed in the comparison for females. These data also closely fit the findings more recently reported by Carrigan and Julian (1966) in which a conformity difference in this same direction was obtained for the sample of males, but no difference as a function of ordinal position was found to exist for the sample of females.

n Achievement

In the 1962 study, a weak but tantalizing relationship between birth order and *n Achievement* was suggested by the data: first-borns had higher *n Achievement* than later borns. In this present study, we can once again examine this relationship. Table 3 presents the means and analysis of variance for these data based on the Edwards Personal Preference Schedule measure of *n Achievement*.⁷ This measure of *n Achievement* indi-

⁷ Although we employed a projective measure of *n Achievement* (French, 1958) in addition to this

cates a significant main effect for ordinal position ($p < .001$) and a significant interaction between subject sex and sibling sex ($p < .005$). Consistent with the finding of the earlier study, in this study, our data indicate that firstborns have higher *n Achievement* than second borns. This difference is best examined in light of the significant interaction between subject sex and sibling sex. This same interactive pattern occurs again in several other analyses.

With regard to these *n Achievement* data, an examination of the means indicates that the two-family configurations in which there is an older brother or a younger sister produce higher *n Achievement* than the comparative configuration having a younger brother or an older sister. Quite simply, these data show that having a younger sister or an older brother is more of a stimulant to the development of high achievement than having a younger brother or an older sister.

The entire picture involving *n Achievement* is amplified further when we examine the data involving the two only-child groups. Each of these groups fits into the general picture in a manner which closely parallels their own sex grouping. The only-child males have an achievement score second only to the M₁F

measure, analysis indicated that only the three-way interaction was significant, thus making interpretations of the data difficult. These data are reported in the complete report of the entire project (cf Hancock, 1965).

TABLE 3
MEAN SCORES FOR *n ACHIEVEMENT* (EDWARDS PERSONAL PREFERENCE SCHEDULE)^a

Item	M	F	M ₁ M	M ₁ F	M ₂ M	M ₂ F	F ₁ M	F ₁ F	[F ₂ M	F ₂ F
\bar{X}	16.07	14.81	15.00	16.72	15.87	14.29	11.93	14.63	13.67	11.90
N	(14)	(16)	(22)	(25)	(23)	(17)	(30)	(32)	(30)	(21)

Analysis of variance

Source of variance	Residual SS	MS	df	F
Ordinal position (A)	4834.475	285.451	1	12.048****
Sex (B)	4568.430	19.406	1	.819
Sex of sibling (C)	4552.486	3.462	1	.146
A × B	4549.996	.972	1	.041
A × C	4550.873	1.849	1	.078
B × C	4729.188	180.164	1	7.604***
A × B × C	4553.047	4.023	1	.169
Error	23.693		192	

^a High number indicates greater *n Achievement*.

*** $p < .005$.

**** $p < .001$.

TABLE 4
MEAN SCORES FOR TEST ANXIETY*

	M	F	M:M	M:F	F:M	F:F	F:M	F:F	F:M	F:F
T	22.08	24.07	24.52	22.80	22.54	22.00	22.00	22.00	22.00	22.00
A	22	18	20	22	20	20	20	20	20	20
Analysis of Variance										
Source of Variation	Between Groups		Within Groups		Total		Error		Total	
Ordinal position (A)	8359.618		139.842		1		3.45*			
Sex (B)	8260.975		11.199		1		.277			
Sex of sibling (C)	8249.884		.108		1		.003			
A × B	8250.111		.000		1		.000			
A × C	8250.231		.455		1		.011			
B × C	8336.166		86.390		1		2.146			
A × B × C	8256.838		7.062		1		.175			
Error	40.440				204					

* High score indicates greater anxiety.

* $p < .10$.

group, while the only-child females' achievement score is similar to that of the F_1F group. This difference between the two groups of only children parallels the sex difference for the entire sample. Although this sex difference is not statistically significant, there is a consistent trend across all comparison groups showing the males to have a higher n Achievement score than the females. Consistent with the interpretation of the interaction effect previously discussed, it is of interest to note that the two female groups which come closest to the male groups in their n Achievement score are the F_1F group, with the apparent positive stimulant of a younger sister, and the only child.

Test Anxiety

One component of n Achievement which Atkinson has discussed involves the avoidant side, referred to as fear of failure (Atkinson, 1958; Atkinson & Litwin, 1960). In his own studies, Atkinson typically measured this end of the achievement dimension by administering a test-anxiety questionnaire. Table 4 presents these data for our sample, involving a shortened version of the Mandler-Sarason Test Anxiety scale (Mandler & Cowan, 1958). Examination of this table indicates that the only effect that approached statistical significance involves ordinal position; second borns have a somewhat higher test-anxiety score than firstborns ($p < .10$). In adding the only-child groups to this picture, we note that the

only-child male scores lowest of all groups on test anxiety, while the only-child females are more similar to the firstborn females than to the second-born females.

Although the Subject Sex × Sibling Sex interaction does not quite reach a significant F value, its patterning is of interest, particularly when considered along with the same effect for the measure of n Achievement. Once again, we note that having a younger sister or an older brother tends to produce less anxiety than having a younger brother or an older sister.

n Affiliation

In addition to the personality measures of n Achievement and test anxiety, which were an important part of the original study which the present study partially replicates, we were able to obtain several other relevant measures. We turn now to an examination of n Affiliation as measured by the Edwards Personal Preference Schedule scale. In theory, one might argue that firstborns are more affiliatively dependent and anxious than later borns (Schachter, 1959). In several studies, firstborns appear to prefer to affiliate with others when made anxious (Schachter, 1959; Wrightsman, 1960), though as Sampson (1965) suggested in his review article the picture with respect to the measured personality trait of affiliation is by no means as clear as the foregoing would suggest.

Table 5 presents the means and the analy-

TABLE 5
MEAN SCORES FOR *n* AFFILIATION (EDWARDS PERSONAL PREFERENCE SCHEDULE)^a

Item	M	F	M ₁ M	M ₁ F	M ₂ M	M ₂ F	F ₁ M	F ₁ F	F ₂ M	F ₂ F
\bar{X}	12.58	15.69	14.46	13.60	13.48	14.38	15.97	14.75	14.13	15.10
<i>N</i>	(14)	(16)	(22)	(25)	(23)	(16)	(30)	(32)	(30)	(20)

Analysis of variance

Source of variance	Residual SS	MS	df	F
Ordinal position (A)	3440.939	20.173	1	1.120
S sex (B)	3449.274	28.508	1	1.583
Sex of sibling (C)	3421.031	.265	1	.015
A × B	3445.371	24.605	1	1.367
A × C	3421.711	.945	1	.053
B × C	3460.248	39.482	1	2.193
A × B × C	3421.106	.340	1	.019
Error	18.004		190	

^a High number indicates greater *n* Affiliation.

sis of variance for our sample of *n* Affiliation. It is plain to see that there were no differences in *n* Affiliation between the eight dyadic groups that composed the complete factorial design. When we bring the only-child subjects into the picture, we obtain a few comparisons which are significant at around the .10 level (e.g., comparisons involving M₀ with F₀ and with the two FF groups).

Although it did not reach a statistically significant *F* value, we turned our attention once again to the interactive pattern that appeared to emerge in our scrutiny of the achievement and anxiety data. As with the preceding two needs, we note that having a younger sister

or an older brother tends to produce a lower *n* Affiliation score than having a younger brother or an older sister.

n Autonomy

Table 6 presents the means and the analysis of variance for our sample on *n* Autonomy. The main effect of ordinal position is highly significant ($p < .001$), with the firstborns scoring higher in autonomy than the second borns. Examination of the only-child groups indicates that they fall at the extremes of their own sex grouping. That is, only-child males have the lowest *n* Autonomy of all male groups, with this difference reaching

TABLE 6
MEAN SCORES FOR *n* AUTONOMY (EDWARDS PERSONAL PREFERENCE SCHEDULE)^a

Item	M ₁ M	M ₁ F	M ₂ M	M ₂ F	F ₁ M	F ₁ F	F ₂ M	F ₂ F
\bar{X}	13.43	13.36	12.00	13.84	10.65	11.47	9.91	10.95
<i>N</i>	(23)	(25)	(24)	(19)	(31)	(36)	(34)	(22)

Analysis of variance

Source of variance	Residual SS	MS	df	F
Ordinal position (A)	4575.717	298.018	1	14.352****
S sex (B)	4293.231	15.532	1	.748
Sex of sibling (C)	4320.007	42.308	1	2.037
A × B	4277.983	.284	1	.014
A × C	4277.732	.033	1	.002
B × C	4292.244	14.545	1	.701
A × B × C	4286.954	9.255	1	.446
Error	20.766		206	

^a High number indicates greater *n* Autonomy.

**** $p < .001$.

significance in comparing only-child males with both groups of firstborn males (.05) and the M_2F group (.05); only-child females score lower in n Autonomy than any of the other female groups, with this difference reaching statistical significance in comparisons involving same-sex sibling pairs (i.e., F_0 versus F_1F , $p < .05$; F_0 versus F_2F , $p < .01$). It is almost as though n Autonomy reflects a rebellion against one's siblings; the only child, lacking such rivalrous sibling relationships, develops a lesser n Autonomy.

DISCUSSION

There are several ways in which one can examine the meaning and implications of these findings. We shall attempt to make some degree of sense out of what are essentially very complex data by first discussing the main effects involving ordinal position, turning only then to the complicating interaction effects.

Our data indicate that in two-sibling families, the single variable of ordinal position accounts for significant differences in n Achievement (.001) and n Autonomy (.001). In interaction with the subject's sex, ordinal position approaches significance in accounting for an overall difference in conformity behavior (.10) while significantly differentiating the conformity of first- and second-born males (.01). Firstborns have higher n Achievement and n Autonomy and tend towards lower test anxiety than second borns, and, if male, conform more in a social influence situation. Adding only children to this picture modifies it slightly. The only-child male is similar to the firstborn in that he is high in n Achievement, low in test anxiety, and high in conformity; he is unlike the firstborn, however, in that he is lower in n Autonomy. The only-child female is similar in many respects to the second born in that she is high in test anxiety, low in n Autonomy, low in conformity, and scores about medium in n Achievement.

Perhaps more complicating to the preceding picture than the addition of the only-child groups is the somewhat puzzling relationship which exists between achievement, autonomy, anxiety, and conformity. If we follow Atkinson's method of dealing with data involving n Achievement (Atkinson, 1958;

Atkinson & Litwin, 1963) one would assume that persons high in n Achievement and low in test anxiety would be more likely to approach achievement than to avoid failure. In other words, these persons are the classically described high need achievers. This description is consistent with the data for the firstborns and the only child males in our sample. As was pointed out in the earlier article (Sampson, 1962), the relationship between n Achievement and conformity is by no means direct or simple. Whether persons who measure high in n Achievement will conform or resist a social influence attempt is in large part a function of the nature of the situation in which that attempt takes place. As both Samelson (1958) and Zajonc and Wahi (1961) have demonstrated, when conformity behavior is perceived as instrumental to achievement, one finds a positive relationship between n Achievement and conformity. In this study, the nature of the instructions which emphasized the importance of doing well and the context of sponsorship which focused on university and school district approval for the project provided ample arousal conditions for n Achievement. Seen in this manner, it is not surprising that the firstborns have higher n Achievement, tend to have lower test anxiety, and, if male, higher conformity. Furthermore, it is not surprising that this arousal has greater effect on the males (witness their greater conformity) than upon the females. Perhaps it is as Carrigan and Julian (1966) suggest, it requires a threat of physical pain to effectively arouse females' conformity or resistant behavior. Their study suggests that even affiliative arousal does not provide the spark that differentially ignites the female conformity response; this study indicates that achievement arousal is equally impotent in this regard.

In the previous article in which sex differences in conformity were related to ordinal position (Sampson, 1962), results not very much unlike those just reported were found. Given our preceding interpretation, perhaps it is better to conclude that males high in n Achievement are more typically inclined to perceive most testing situations (i.e., most experimental situations) as involving some test of their achievement, and thus they re-

spond accordingly by conforming when by so doing in a given situation they are also brought closer to achievement. Females, on the other hand, are less likely to respond in this manner, for they are unlikely to perceive such situations as a challenge or a test of their achievement. Our conclusion therefore from this and the previous study that there is a Sex \times Ordinal Position difference in conformity behavior may very well be more a function of differential arousal than a difference in the abstract notion of conformity. That is, for females we really may not yet know what the relation is between birth order and conformity because we have not discovered, in any systematic manner, the conditions which arouse in them the relevant intervening psychological processes. For males, on the other hand, achievement arousal appears to be meaningful intervening process.

If the ordinal position effects which we have examined, including those involving personality variables and conformity behavior, are to have any psychological meaning, this meaning must exist in the structure of the family and in the particular kinds of experiences this structure provides for the child. As several of the preceding analyses suggested, ordinal position is but one determinant of family structure; the interaction between the sex of the subject and the sex of his sibling provides another significant focal point for structural differentiation.

Although the understanding of these intra-familial structural differentiations and the complex possibilities they offer is beyond the appropriate scope of our present data, we would suggest that the Parsonian theory of family-role differentiation may be especially valuable to further work in this area (Parsons, 1955). According to this framework, one's sex places one along an instrumental-expressive role dimension—fathers and sons fill instrumental roles, mothers and daughters, expressive—while one's generation and ordinal position determine placement along a power axis—parents are power superior to their children; first are power superior to second. With regard to intrafamilial dynamics, the model assumes tendencies towards a balanced family-role configuration in which all four basic status roles are fulfilled: instru-

mental and expressive, power superior and inferior. Such tendencies towards balance in the family-role system can in turn create certain key incongruities for an individual family member between his role within the family and his role within the broader culture. Thus, for example, in a family with two children of the same sex, balance within the family-role structure is achieved by allocating a culturally inappropriate sex role to one of the pair. Culturally inappropriate sex roles can also occur when, for example, the pair is cross sex, but the firstborn is a female and thus is power superior. In this instance, the female's superior position and the male's inferior position in the family are incongruent with broader cultural expectations for their sex-appropriate behavior. One would expect that these hypothesized changes within the role system would in turn be reflected in measurable personality variables which tapped sex-relevant dimensions.

Although these data were not collected with an eye to testing any part of the Parsonian model, it will be of some interest and value to examine briefly two instances where the framework helps to organize some of these present data. The family configuration involving an older female with a younger sister (F_1F and F_2F) presents a picture in which the older female is lower in anxiety and affiliation but higher in achievement and autonomy than her younger sister. This shift on the part of the older female with a younger sister toward the more masculine instrumental pattern (i.e., higher achievement and autonomy) while the younger sister retains the more typically feminine expressive functions (higher anxiety and affiliation) is consistent with the model.

The cross-sex configuration involving an older male with a younger sister (M_1F and F_2M) provides another instance of confirmation for the model. This arrangement is presumed to be congruent, in that roles within the family fit with sex-appropriate roles expected within the broader culture. For this pattern, we find the older male is lower in anxiety and affiliation and higher in achievement and autonomy than his younger sister. That is, we find a typically masculine pattern for the older male and a typically feminine pattern for the younger female. In this in-

stance there has been no crossover in roles (and personality) in order to achieve a balanced family configuration as we encountered in the first example.

We present these last examples more as suggestive of a valuable direction for further study than as well-defined confirmation for a developed theoretical framework. It seems reasonable to expect that a family role analysis will provide one useful manner of giving meaning to the too often psychologically meaningless variable of ordinal position. Such an analysis should be especially helpful in systematically examining and understanding the complex interactions between ordinal position and other family-structure variables such as sibling sex. Should this line of investigation be pursued, it would necessitate examining different positions and sex roles within the same family. Perhaps then one could bring together into a more meaningful theory the often confusing collage of literature on ordinal position.

REFERENCES

- ATKINSON, J. W. (Ed.) *Motives in fantasy, action, and society*. Princeton, N. J.: Van Nostrand, 1958.
- ATKINSON, J. W., & LITWIN, G. H. Achievement motive and test anxiety conceived as motive to approach success and motive to avoid failure. *Journal of Abnormal and Social Psychology*, 1960, 60, 52-63.
- BECKER, S. W., LERNER, M. J., & CARROLL, J. Conformity as a function of birth order and type of group pressure: A verification. *Journal of Personality and Social Psychology*, 1966, 3, 242-244.
- CAPRA, P. C., & DITTES, J. E. Birth order as a selective factor among volunteer subjects. *Journal of Abnormal and Social Psychology*, 1962, 64, 302.
- CARRIGAN, W. C., & JULIAN, J. W. Sex and birth-order differences in conformity as a function of need affiliation arousal. *Journal of Personality and Social Psychology*, 1966, 3, 479-483.
- DEUTSCH, M., & GERARD, H. B. A study of normative and informational social influences upon individual judgment. *Journal of Abnormal and Social Psychology*, 1955, 51, 629-636.
- EDWARDS, A. L. *Edwards Personal Preference Schedule*. New York: Psychological Corporation, 1953.
- FRENCH, E. G. Development of a measure of complex motivation. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton, N. J.: Van Nostrand, 1958. Pp. 242-248.
- HARRIS, E. J. An empirical investigation of the functioning of ordinal position, sex, and sex of sibling in socialization, personality, and family behavior among adolescents in two and three-child families. Unpublished doctoral dissertation, University of California, Berkeley, 1967.
- HARRIS, W. L. *Children in the family*. New York: Holt, Rinehart & Winston, 1963.
- MORRIS, G. C., & CROW, J. E. The family position variable. *Journal of Consulting Psychology*, 1962, 22, 25-29.
- POWELL, T. Family structure and development of the child. In E. Fennell & E. F. Hulse (Eds.), *Family structure and development*. Glencoe, Ill.: Free Press, 1965. Pp. 141-151.
- SACHS, E. E. Birth order, need achievement, and affiliation motives for achieving success in two conditions of conflict with a partner. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton, N. J.: Van Nostrand, 1958. Pp. 314-321.
- SACHS, E. E. Birth order, need achievement, and affiliation motives. *Journal of Abnormal and Social Psychology*, 1960, 64, 156-159.
- SACHS, E. E. The study of ordinal position. *Approaches and outcomes*. In B. Maher (Ed.), *Progress in experimental personality research*. Vol. 2. New York: Academic Press, 1965. Pp. 175-185.
- SCHNEIDER, S. *The psychology of children*. Stanford: Stanford University Press, 1967.
- SCHULTE, H. *The analysis of variance*. New York: Wiley, 1969.
- SHEDDEN, P. Birth order of volunteers for anxiety description. *Journal of Abnormal and Social Psychology*, 1961, 68, 145-147.
- WAGNER, W. L., MURPHY, M., & FARR, K. *Social class in America*. Chicago: Science Research Associates, 1962.
- WAGNER, J. R. Birth order and social behavior. *Psychological Bulletin*, 1959, 65, 38-49.
- WEISS, J. H., WOLF, A., & WAGNER, R. G. Birth order, treatment conditions, and preferences for participation in group versus nongroup experiments. Paper read at American Psychological Association, Philadelphia, August 1966.
- WINTERSTROM, M. R. The relation of need for achievement to learning experiences in intelligence and mastery. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton, N. J.: Van Nostrand, 1958. Pp. 445-478.
- WRIGHTSMAN, L. Effects of waiting with others on changes in level of felt anxiety. *Journal of Abnormal and Social Psychology*, 1960, 61, 216-222.
- ZATONE, R. B., & WAHL, N. K. Conformity and need achievement under cross-cultural norm conflict. *Human Relations*, 1961, 14, 241-250.

(Received June 7, 1966)

TRAINING CHARITY IN CHILDREN¹

ELIZABETH MIDLARSKY AND JAMES H. BRYAN

Northwestern University

160 1st- to 4th-grade girls were given 1 of 10 training programs designed to produce variations in altruistic behaviors in 2 distinct donation situations. Children sacrificed the obtaining of M & M candies when they had a warm relationship with E who made explicit her pleasure at such sacrifice. Neither a warm relationship nor explicit statements of joy by E alone were more effective than no training in eliciting such charitable behavior. Ss' subsequent anonymous donations were highly correlated with previous training programs. It was concluded that positive interpersonal relationships and explicit statements of pleasure by the socializing agent can provide the basis for the internalization of the norm of self-sacrifice.

While the study of the preconditions and correlates of antisocial acts has had a long and honorable history, only recently has attention been directed toward the scientific study of altruistic responses. Concern has been evidenced, both at the empirical and theoretical level, about those behaviors in which the person makes some sacrifice for apparently no material or personal gains. Campbell (1965), arguing against hedonistic models of man, has proposed that altruistic behavior may actually find its roots in biological evolution and that there is little reason to assume that nonegocentric motives are of less importance to survival than self-centered ones. The recent experiments of Berkowitz and his colleagues have demonstrated that college students will show more helping behaviors toward dependent than independent others, even though possibilities of reward are remote (Berkowitz & Daniels, 1963; Berkowitz, Klanderman, & Harris, 1964). Their general finding that dependency elicits helping behaviors in spite of the absence of externally administered reinforcement has been replicated by Schopler and Bateson (1965).

Some attention has been given to sharing behaviors in children. Ugurel-Semin (1952) found that the sharing of peanuts by young children was positively correlated with age

and family size, the former correlation being subsequently replicated by Handlon and Gross (1959). Unfortunately, neither study controlled for the possibility that changes in preferences accounted for the increased sharing. Experimental studies of self-sacrificing responses by children are, however, quite rare. Fischer (1963) found that giving marbles to an unknown peer was facilitated when the subject was reinforced with bubble gum. He thus argued that sharing behavior is essentially giving up one reward for the promise of another. Other logical possibilities were untestable by virtue of his experimental design. Evidence exists that social models are influential in producing charitable responses in children. White and Rosenhan (1966) demonstrated that the observation of a charitable other will facilitate donations of gift certificates to a fictitious orphanage.

Aronfreed and Paskal (1965) have provided data to support their suggestion that sacrificing behavior can be enhanced by attaching, through contiguous association, positive affect in the donor with expressions of joy in the receiver. Such expressions on the part of others become stimuli for the arousal of positive affect in the donor. These investigators found that pairing expressions of joy by the experimenter with the hugging of the subject elicited more subsequent sacrificing of M & M candies for such joyous responses than either the administration of the hug or joy response alone.

The present investigation was designed to assess the role of affective arousal and ex-

¹ This experiment is based on a thesis submitted by the senior author in partial fulfillment of the requirements for the MA degree at Northwestern University, 1966. Thanks are due to the many helpful suggestions of Winfred F. Hill, David Rosenhan, and Edward L. Hoffman.

pression of joy upon subsequent donation behavior. Although Aronfreed and Paskal suggested that the patterning of the expressive cues and the hug response was important in the development of self-sacrifice, no attempt was made to test this assumption. Additionally, the subject's behavior was always visible to the experimenter and, since subjects in the experimental groups received many more cues as to the experimenter's desires, their finding might reflect conformity to an external (i.e., the experimenter) rather than to an internal (i.e., the internalized norm) source. The current study was addressed to these issues, as both patterning and degree of expressive and affectional responses varied, and the charitable responses were measured under conditions of donor anonymity.

METHOD

Subjects

Subjects were 160 Caucasian girls, randomly selected from the first through fourth grades of a local public elementary school. The majority of children were from middle-class families. Children were randomly assigned to 1 of the 10 treatment conditions, the only restriction being that four children from each grade were represented in each condition. The experimenter was a 25-year-old female graduate student.

Apparatus

The apparatus was similar to that described by Aronfreed and Paskal. It consisted of a boxlike enclosed instrument case, approximately 9 inches high, 27 inches wide, and 15 inches deep. At the center of the case was a chute through which M & Ms were dispensed. Above the chute lip was a circular red light, 1 inch in diameter. Protruding from the front of the case were two levers, each 8 inches from either side of the center. The right-hand lever was programmed to activate the red light; the left-hand lever dispensed one piece of candy at each pull. Both were programmed to 60% probability in a variable-ratio schedule. A second red light was placed on the right-hand side of the case, out of the subject's sight. Next to the light was a double-pole, single-throw toggle switch, the operation of which could cause the red light on the side of the case to be activated in lieu of the front red light.

Upon entering the experimental room (located within the school) each child was informed that the experimenter was testing a big box and that she was there to help her in this regard.

Prior to the training, each subject was asked to indicate whether she (a) really loves M & Ms, (b) likes M & Ms a bit, (c) does not care about

M & Ms, or (d) really does not like M & Ms. Each subject, after being told that she would be allowed to keep the M & Ms, was given the following instructions:

You see, this box has two handles, one here and one here [while touching each], and each does something different when we use it. If I were to use this one then sometimes, just sometimes a red light will go on. If I were to use the other, then sometimes an M & M candy will come down through the chute and into your cup. If you see an M & M, you are to leave it in your cup until we are all through. Now, watch me first and I will show you how the box works.

In the training situation, the experimenter gave 20 demonstration trials in the use of the apparatus. The order in which the levers were described to the subject, as well as the order in which they were demonstrated, was systematically varied. For half of the subjects, the demonstration began with the left-hand lever, and for half it began with the right-hand one. The demonstration consisted of four blocks of five trials each. During the first five trials the experimenter demonstrated the use of the right- or left-hand lever. In the second sequence of five trials, she demonstrated whichever lever had not been used previously. The third and fourth blocks of trials repeated the order used in the first and second. Prior to each block of trials, the experimenter stated, "Now I'll try this one," or "Now I'll use this handle a few trials." Whenever the red light was activated, the experimenter reacted as dictated by the particular training condition. During this demonstration, the five training programs were initiated. Subjects were exposed to one of the following conditions:

1. The E-H condition in which the onset of the red light elicited the expressive cue of "Oh good, I see the red light," or "Oh, there's the red light again," or "I see the light," followed by a gentle hug of the subject. The expressive cues were said in such a manner as to indicate the experimenter's joy. This treatment employed forward conditioning procedures.

2. The H-E condition in which the same procedures were followed as indicated under the E-H condition except that the hug preceded and was distinct from the expressive cues. Backward conditioning procedures were then employed.

3. The H condition where the onset of the red light cued the experimenter's hugging the subject.

4. The E condition where only the expressive cues were given after the red light flashed.

5. The O condition where neither the expressive cue nor the hug followed the onset of the light.

Following the training trials, the experimenter explained to the subject that she would now be allowed to play with the box by herself while the experimenter sat on the side and watched. The subject was told that the red light in front of her would be turned off "because we don't want to use it all up," but that there was another light which

could be seen from the experimenter's seat. The subject was informed that she was to be rewarded that she could earn candy that she could keep for herself if she took the hand given and the hand was not seen. She was to be rewarded that she could earn candy if she took the hand. There were 10 trials with a 10-second interval of 1 second in between. Half of the subjects of each of the training groups received the expressive cues from the experimenter whenever the red light was flashed (E groups); the remaining subjects did not (O groups). During the test trials, the experimenter sat at the end of the table with her gaze directed at the apparatus. The experimenter, when administering expressive cues during the test trials, refrained from looking at the subject.

Following termination of the last trial, the experimenter gave the subject all the additional number of M & Ms and simultaneously thanked her for her help. While the fact that the subject was given a reward was obvious, the procedure was such as to obscure the exact number of M & Ms given. The purpose of this additional reward was to equate the subjects within the various training groups for "wealth" prior to donation task (Task II). At the termination of Task I, each subject then had 50 M & M candies.

For Task II, each subject was asked to donate candy to "needy children, whose parents can't afford to buy them any candy" by placing some of her own candy through the slot in a half-filled donation box, whose contents were invisible to the subject. In order to create anonymity for the donor, the subject was led behind a screen in the experimental room, where the donation box was located under the sign making the above-mentioned appeal. Each subject was told that the size of donation was her decision, and that she need not donate any candy if she did not choose to do so. The subject was informed that no one would ever know what she had decided to do and that the experimenter would wait for her outside of the experimental room. Since all the M & Ms in the donation box were brown (unknown to the subject), and subjects received other colored candies, the amount donated by each subject was easily obtained. The subject was unable to see the contents of the box. Crepe paper was put on the box in such a manner as to detect any incidents of the subject removing the M & Ms from the container. No such incidents occurred. After completing Task II, each subject was escorted to his classroom.

RESULTS

The number of occasions the subject pressed the lever *not* associated with M & M candies (self-sacrifice response) served as the dependent variable in Task I, while the number of M & Ms donated to needy children served as the measure on Task II. Table 1 presents the means and standard

deviations of performance according to the training and test conditions on both Task I and Task II. Since preference scores for M & M candies were not related to performance on either Task I or Task II ($r_{01} = .10$, $r_{02} = .18$, respectively), they were not employed as an adjustor variable in the subsequent analysis.

Table 2 presents the analysis of variance of the results of Task I. In the between-subject comparisons, the main effects of type of training and of subsequent expressive behavior on the part of the experimenter were significant, as were the interactions of training with age, and training with expressive behavior. The η^2 's for the main effect were .41 and .27 and for the interactions, .28 and .22, respectively. Thirty-seven percent of the

TABLE 1

MEANS AND STANDARD DEVIATIONS OF SELF-SACRIFICING RESPONSES (TASK I) AND (TASK II) DONATIONS

	Task I	Task II
HE		
E		
\bar{X}	26.38	24.75
SD	7.94	9.25
O		
\bar{X}	17.63	15.31
SD	5.39	8.32
EH		
E		
\bar{X}	24.63	24.69
SD	3.85	8.58
O		
\bar{X}	17.88	13.88
SD	3.43	6.04
E		
E		
\bar{X}	16.75	16.63
SD	6.05	10.31
O		
\bar{X}	14.81	15.50
SD	3.28	6.18
H		
E		
\bar{X}	16.88	13.06
SD	4.69	6.68
O		
\bar{X}	14.88	14.25
SD	5.99	12.14
O		
E		
\bar{X}	15.25	16.00
SD	4.50	7.14
O		
\bar{X}	14.56	14.94
SD	5.79	12.48

TABLE 2

ANALYSIS OF VARIANCE SUMMARY: TASK I

Source of variation	df	SS	MS	F ratio
Between Ss				
Training cues (C)	4	360.93	90.23	14.56**
Grade (G)	3	43.58	14.53	2.34
Expression (E)	1	159.00	159.00	25.66**
C × G	12	171.38	14.28	2.30*
C × E	4	100.57	25.14	4.06*
G × E	3	42.62	14.21	2.29
C × E × G	12	45.97	3.83	
Error between	120	743.69	6.19	
Within Ss				
Blocks of trials (B)	3	11.57	3.86	4.41**
B × C	12	35.83	2.99	3.41**
B × G	9	18.71	2.08	2.38**
E × B	3	6.15	2.05	2.34
C × G × B	36	54.61	1.52	1.73*
C × E × B	12	10.87	.91	1.03
G × E × B	9	12.00	1.33	1.52
C × G × E × B	36	27.44	.76	
Error within	360	315.06	.88	
Total	639	2159.99		

* $p < .05$.
 ** $p < .01$.

total variance was associated with the main effects and interactions. The within-subject comparisons indicated that significant changes were occurring across blocks of trials, and that the effects of trials were interacting with types of training and grade of the subject. In addition, a second-order interaction was significant, with the variables training, grade level, and trial blocks involved.

All differences among means of selected subgroups were analyzed by the Newman-Keuls procedure (Winer, 1962) with α set at .05. E-H and H-E treatments were more effective in inducing the sacrificing of M & Ms than the other treatment conditions. Under those conditions where the expressive cue was not subsequently introduced into the test trials, both E-H and H-E training programs were significantly more influential than any of the control groups, while the latter groups showed no significant differences among themselves. Under conditions where the expressive cue was introduced into the testing trials, the H-E group was significantly more affected by the training than the E-H, but both were more affected than their controls. Again the control groups did not differ. It is also apparent from Table 1 that there is an interaction of training with expressive cues

on Task I, with E-H and H-E programs being markedly affected by the subsequent introduction of "joy" responses during the test trials.

The significant Grade × Training interaction is shown in Table 3. Analysis of subgroup differences was accomplished by applying the Newman-Keuls to each variable (either grade or training) while holding the second variable constant. While no differences were obtained between grades under the H-E and H conditions, second graders donated significantly more than all others under the E-H condition, and first graders donated significantly less than all others under the E condition.

Within-subject comparisons indicated that while there were significant decreases in the self-sacrifice response over the course of trials, the trials variable showed significant interactions with types of training and the grade of the subject. These interactions are depicted in Figures 1 and 2.

It is apparent from Figure 1 that the two primary training programs (E-H, H-E) tended to sustain the self-sacrifice response over the course of the trials, while the control groups showed the expected seeking of M & M candies. Analysis revealed no significant changes occurred over blocks of trials in the E-H and H-E groups. However, significant decrements were obtained from Block 1 to Block 3 in the E group, from Block 1 to Block 4 of the O group, and from Block 1 to other blocks in the H condition.

Figure 2 demonstrates that while second-grade subjects increased self-sacrifice responses over trials, subjects at other grade levels did not. Because of the difficulty, the

TABLE 3

MEAN SCORES OF DONATIONS DEPICTING THE TRAINING × GRADES INTERACTION ON TASK I

Condition	Grades			
	1	2	3	4
HE	23.00	21.75	20.25	23.00
EH	20.38	24.63	19.25	20.75
E	9.75	18.38	18.88	16.13
H	15.38	16.25	15.00	17.12
O	12.88	11.63	17.00	18.25

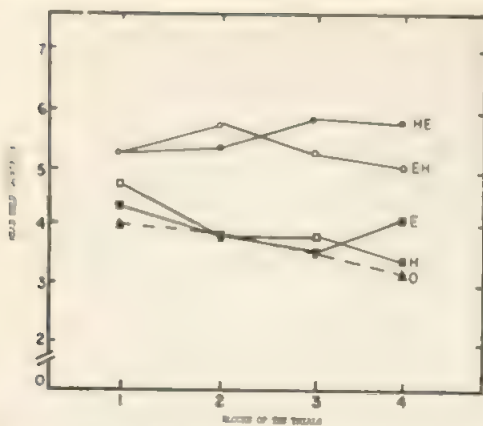


FIG. 1. Curves for performance over blocks of 10 trials for five training groups on Task I.

second-order interaction of Training \times Grade \times Trials will not be interpreted.

There was considerable transfer from Task I to Task II. Pearson product-moment correlations between these two types of sacrificing equaled .67. Furthermore, the median within-cell correlation of Tasks I and II for the 10 groups also equaled .67, with the range between .10 (the H-E group) to .82 (EH-E group). The remaining within-cell correlation across the two tasks exceeded .64 with the exception of the O-E group; the latter group's correlation between Tasks I and II equaled .43. All but the two lowest correlations were significant ($p < .05$).

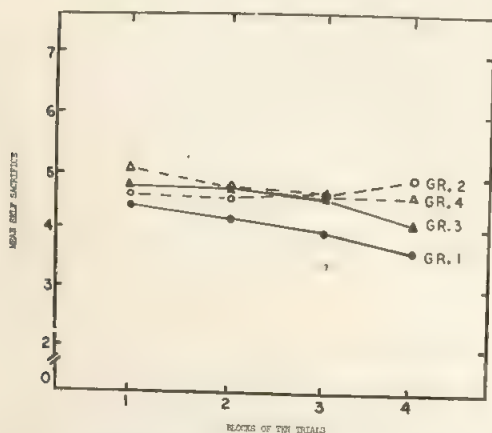


FIG. 2. Curves for performance over blocks of 10 trials for four grade levels on Task I.

TABLE 4
ANALYSIS OF VARIANCE SUMMARY: TASK II

Source of variation	df	SS	MS	F ratio
Training cues (C)	4	919.90	229.98	3.30**
Grades (G)	3	1479.05	493.02	7.20**
Expression (E)	1	722.50	722.50	10.55**
C \times G	12	1534.45	127.87	1.87*
C \times E	4	955.75	238.94	3.49**
G \times E	3	4.25	1.42	
C \times G \times E	12	581.00	48.42	
Within groups	120	8221.50	68.51	
Total	159	14418.40		

Note.—The remaining interactions were not significant.

* $p < .05$.

** $p < .01$.

The results of the analysis of variance for Task II are presented in Table 4.

The main effects of training, grade, and expressive cues are significant and are associated with 6%, 10%, and 5%, respectively, of the total variance, while the interactions of Training \times Expressive Cues and Training \times Grade are significant, accounting for 11% and 7% of the total variance, respectively.

Table 1 depicts the interaction of the training programs with expressive cues in eliciting donations to "needy children" (Task II). As predicted, the combination of subsequent expressive cues with either the E-H or H-E training programs elicited the greatest degree of charity. While there were no significant differences attributable to training technique upon donations in Task II in the absence of subsequent expressive cues, the training programs in combination with subsequent "joy" responses yielded statistically significant differences. Both the E-H and the H-E groups who subsequently received expressive cues donated significantly more than

TABLE 5
MEAN SCORES OF DONATIONS ON TASK II DEPICTING
THE TRAINING \times GRADES INTERACTION

Condition	Grades			
	1	2	3	4
HE	20.25	15.88	20.75	23.25
EH	18.38	21.13	18.25	19.38
E	6.25	18.38	22.38	17.25
H	8.63	14.75	15.25	16.00
O	7.75	10.75	23.50	19.88

their controls; the latter groups showed no significant differences among themselves.

The means which reflect the significant interaction of training with grades are shown in Table 5. Analyzing subgroup means, across ages, with training conditions held constant, few subgroup differences were found. First graders donated significantly less than all other groups under the E training paradigm, while both first and second graders were considerably less charitable than third and fourth graders under the no training condition. When grades were held constant and treatments compared, significant differences were obtained only in the case of the first graders. First-grade subjects in the H-E and E-H groups were significantly more altruistic than those in the other treatment conditions.

DISCUSSION

The results clearly indicate that charitable behavior under conditions of donor anonymity was significantly related to training program and the age of the child. Self-sacrificing behavior had occurred under conditions where little material or social reward was apparent.

The mechanisms underlying such behavior, however, are not obvious. The suggestion that the expressive cues become reinforcers through contiguous association with the subject's positive affect was not supported. While there was suggestive evidence that expressive cues introduced during the test trials on Task I were reinforcing to the subject (Cue \times Trial interaction = $p < .10$), the lack of a significant second-order interaction of Trials \times Expressive Cues \times Training makes untenable the notion that the training programs produced the reinforcement properties of these cues. A secondary reinforcer was not established through the pairing of hugs and expressive cues. This finding is similar to that of Aronfreed and Paskal's that subjects were no more inclined to press for the red light during the last than the first half of the 30 test trials. Furthermore, contrary to the implications of Aronfreed and Paskal's position, the patterning of the cues and affect was not found important.

Several hypotheses can be forwarded to account for the Training \times Expressive Cue in-

teraction. It might be argued that the effect of the HE-E and EH-E training programs was due to the punitive effect of the experimenter's silence during the test trials for those subjects with whom she had previously had a warm relationship (i.e., subjects in the E-H-O and H-E-O groups). One might expect, however, that if this hypothesis were true the highly punished would be less inclined than the nonpunished to attempt to elicit the red light. The performance curves of the trials on Task I, however, do not support this contention; those subjects who had a demonstrative experimenter who subsequently withdrew from the interaction behaved no differently than subjects who had a less demonstrative and silent trainer.

It might be hypothesized that subjects in the E-H and H-E training programs who subsequently received the expressive cues during the test trials had a greater number of suggestions as to the experimenter's expectations of their performance. While a hypothesis concerning "demand characteristics" cannot be ruled out, the notion that the child was trained *only* as to which lever the experimenter wanted her to press seems unlikely considering the degree of transfer across Tasks I and II.

A related interpretation of the findings on Task I is based upon the perceived dependency of the experimenter. The implication of experimenter "need" is not a subtle one when, during training, the experimenter both hugs and makes explicit her delight and, subsequently during the test trials, reaffirms these contingencies. While data concerning the effects of perceived dependency upon subsequent helping behaviors in children are not available, a good deal of evidence exists that perceived dependency will elicit helping behaviors by adults (Berkowitz & Daniels, 1963; Berkowitz et al., 1964; Schopler & Bateson, 1965).

It also should be noted that the impact of the training programs on Task I differed according to the grade of the subjects. Apparently, older children are more willing to sacrifice M & Ms than younger; thus the training programs appeared particularly effective with the younger subjects. While it

'is possible that our questionnaire regarding preferences for M & Ms was insensitive, the questionnaire results tend to make implausible the hypothesis that differential preferences account for the variation in either self-sacrifice responses or donations exhibited by the groups. There were virtually no differences between subjects of varying grades regarding their tastes for M & Ms. While the two previous studies (Handlon & Gross, 1959; Ugurel-Semin, 1952) concerned with developmental correlates of sharing behaviors in children did not control for the possibility of preference differences in the varying age groups, both were consistent in demonstrating that sharing behavior increases with age. The present study replicates these findings and suggests that they reflect the increased probability that the children will have learned the general value of charity.

From the results of Task I alone, however, it is impossible to determine whether the children's behavior stemmed from cognitive cues or simply reflected the development of a lever-press response.

Using the results from Task II, both the significant main effect of expressive cues and training as well as the significant interaction between these variables indicate that the training employed during Task I had significant impact upon behavior on Task II. The main effect of age as well as the interaction of Training \times Age indicate that such training had more impact upon subjects from the lower than the higher school grades.

The marked correlation across conditions in which very different motoric behaviors were involved suggests that such behavior was mediated, and that such mediation was highly associated with the various training and test conditions during Task I. These particular programs were differentially effective in eliciting the subject's sacrifice of a prized object for the sake of anonymous other. An internalized norm appeared to have been developed.

The present study, however, does not allow for unequivocal statements as to the mechanisms underlying the elicitation of such charitable behavior. Several hypotheses are possible. First, in equating subjects for M & M

wealth after Task I, subjects' previous behavior may have been reinforced. In the case of the control subjects, this reinforcement would have the effect of maintaining M & M acquisition, while the converse would hold true for the experimental subjects. Furthermore, it is of course possible that since self-sacrificing subjects were directly rewarded by the experimenter, anticipations of further reward after Task II may have been developed by the subjects. Insofar as the experimenter was outside of the room waiting to escort the subject back to the classroom, such anticipations were reasonable.

A second hypothesis to account for the transfer is that the training programs acted to modify the incentive values attached to the candies. The experimenter provided a model who virtually ignored the M & Ms. Since there is evidence that a model who is warm, kind, and nurturant is more likely to be imitated than a less demonstrative or hostile model (Bandura & Huston, 1961), it would be reasonable to assume that incentive values indicated by the experimenter might also be adopted by the observer. The experimenter's restatements of those contingencies affecting her joy under conditions where the experimenter is liked might well modify the incentive value of an otherwise prized object.

Whether modeling or direct reinforcement underlies the internalization of a norm of charity, the present results are that charitable behavior can be significantly altered through the demonstrativeness of the socializing agent.

REFERENCES

- ARONFREED, J., & PASKAL, V. Altruism, empathy and the conditioning of positive affect. Unpublished manuscript, University of Pennsylvania, 1965.
- BANDURA, A., & HUSTON, A. C. Identification as a process of incidental learning. *Journal of Abnormal and Social Psychology*, 1961, 63, 311-318.
- BERKOWITZ, L., & DANIELS, L. Responsibility and dependency. *Journal of Abnormal and Social Psychology*, 1963, 66, 429-436.
- BERKOWITZ, L., KLANDERMAN, S. B., & HARRIS, R. Effects of experimenter awareness and sex of subject and experimenter on reactions to dependency relationship. *Sociometry*, 1964, 27, 327-337.
- CAMPBELL, D. T. Ethnocentric and other altruistic motives. In D. Levine (Ed.), *Nebraska symposium on motivation*: 1965. Lincoln: University of Nebraska Press, 1965. Pp. 283-311.

- FISCHER, W. F. Sharing in preschool children as a function of amount and type of reinforcement. *Genetic Psychology Monographs*, 1963, **68**, 215-245.
- HANDLON, B. J., & GROSS, P. The development of sharing behavior. *Journal of Abnormal and Social Psychology*, 1959, **59**, 425-428.
- SCHOPLER, J., & BATTSON, N. The power of dependence. *Journal of Personality and Social Psychology*, 1965, **2**, 247-254.
- UGUREL-SEMIN, R. Moral behavior and moral judgment of children. *Journal of Abnormal and Social Psychology*, 1952, **47**, 468-474.
- WHITE, G. M., & ROSENHAN, D. L. Elicitation of charitable behavior in children. *American Psychologist*, 1966, **21**, 628. (Abstract)
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received June 13, 1966)

THE 1964 PRESIDENTIAL ELECTION AND CURVES OF INFORMATION SEEKING AND AVOIDING

RAMON J. RHINE

University of California, Riverside

The 1964 presidential election was used as a vehicle to investigate curves of information seeking and avoiding predicted by dissonance theory. Subjects, committed to 1 or the other of the candidates and clearly convinced of Goldwater as more conservative than Johnson were confronted with several degrees of dissonant contradiction about the candidates' relative conservatism. They were offered a choice among political pamphlets whose titles were modeled after items from a political economic conservatism scale. The degree to which the several groups sought or avoided these titles approximated the general inverted U curve of combined seeking and avoiding expected from dissonance theory. The points on this curve differed by analysis of variance. An adjustment in the relative steepness of the predicted avoid and seek curves is suggested by the data to explain differences found for avoidance but not for seeking. A control group received the same treatment as the experimental groups, except that the contradictions used to introduce dissonance were not experienced. In combined seeking and avoiding, the control group exhibited significantly greater information selectivity than the experimental groups.

Festinger (1957, p. 130) specified curves predicting the relation between dissonance and information seeking and avoiding. If both seek and avoid curves are plotted above an abscissa representing increasing degrees of dissonance, they are each approximated by an inverted U. If seeking of consonant information and avoiding of dissonant information are combined, the inverted U again obtains. The purpose of this study was to see if the predicted shape of the curves would be produced, using a potent issue to create several levels of dissonance.

The relation of information seeking and avoiding to dissonance is one of Festinger's (1957, p. 3) two basic hypotheses, and experiments on the topic are numerous (Adams, 1961; Brock, 1965; Brodbeck, 1956; Canon, 1964; Ehrlich, Guttman, Schönbach, & Mills, 1957; Feather, 1962, 1963; Freedman, 1965a, 1965b; Freedman & Sears, 1963; Jecker, 1964; Maccoby, Maccoby, Romney, & Adams, 1961; Mills, 1965a, 1965b, 1965c; Mills, Aronson, & Robinson, 1959; Mills & Ross, 1964; Rosen, 1961; Sears, 1965, 1966; Sears & Freedman, 1963, 1965). Some of these studies have been criticized by Cha-

panis and Chapanis (1964), Festinger (1964), and Steiner (1962). Taken as a whole, Freedman and Sears (1965) have found them lacking consistent support for dissonance theory. The design of these studies precludes convincing support for dissonance theory because they commonly use two dissonance levels to test a prediction which takes the form of an inverted U. Festinger's (1957) study of gambling behavior and its successful repetition (Cohen, Brehm, & Latané, 1959) are the only studies in which data on information selectivity were collected for several levels of dissonance. Therefore, no other existing studies test the total prediction indicated by the curves. Because the gambling situation was complex, the data were not interpreted as a direct reproduction of the predicted curves. Furthermore, since the meaning of the gambling studies is open to question, due to the possibility of an alternative explanation and the complexity of the situation (Chapanis & Chapanis, 1964), further evidence is desirable.

The present study is an attempt to capitalize on a potent and realistic situation, the 1964 election contest between President Johnson and Senator Goldwater. Freedman and Sears (1963) found evidence for selective exposure in an election environment. Dissonance between subjects' commitment to a

¹ Thanks are due to Larry Ogilvie for help in collecting the data and analyzing the results. This research was supported by an intramural grant from the University of California.

candidate and the opposition's propaganda presumably occurred in the Freedman and Sears study. However, as these authors pointed out, the existence of selective exposure does not itself support dissonance theory. In the 1964 election, the Goldwater camp took pains to identify itself publicly as conservative and to contrast its philosophy with that of the Democrats. Although general differences in philosophy may be accepted by voters, specific conservative or liberal positions taken by the candidates are hard to remember and identify. Subjects clearly accepting the existence of the general difference in philosophy, when faced with authoritative contradictions in specifics, should experience dissonance. Then, according to dissonance theory, seeking or avoiding information as a function of dissonance should take the general form of an inverted U.

METHOD

Subjects and Materials

A 10-part questionnaire was administered to 151 undergraduates. The order of presentation of the 10 parts was as follows:² (a) 10 of the most discriminating items of the anti-Semitism scale of *The Authoritarian Personality* (Adorno, Frenkel-Brunswick, Levinson, & Sanford, 1950), (b) two questions, one asking how the subject would vote in the 1964 presidential election between President Johnson and Senator Goldwater and the second asking if the subject favored his choice strongly over the opponent, or somewhat, or slightly, (c) a question asking the subject to say if he does or does not believe news reports indicating that Goldwater is partly Jewish, (d) a request for reasons for the answer to the previous question, (e) nine items from the Political-Economic-Conservatism (PEC) scale (Adorno et al., 1950) for which the subject was asked to give his own opinion, (f) eight different items from the PEC scale for which the subject was required to respond as he thought Johnson would, (g) a repetition of these same eight items with the subject responding as he thought Goldwater would, (h) a request to indicate which of six further PEC items, three from the liberal side and three from the conservative, were themes publicly supported by Johnson or Goldwater, (i) repetition, after a suggestion that the study was over, of the same six items with a statement saying which of the candi-

dates supported and which opposed each theme, and, finally, (j) 12 randomly ordered pamphlet titles with the subject asked to choose the 3 he would most like to have and the 3 he least wanted. The PEC items were used to select subjects who clearly conceived of Goldwater as more conservative than Johnson, and the PEC themes (parts h and i) were employed to introduce assertions contradicting the subject's beliefs about the candidates' conservatism.

Part j of the questionnaire (pamphlets) provides the data of the experiment. The instructions accompanying it were as follows:

As a small expression of appreciation for your cooperation, it may be possible to make available authentic pamphlets summarizing the candidates' official views on many of the issues mentioned in the questionnaire. If the cost is not too great, we will try to obtain copies for you to keep. Look over all the pamphlets listed below and indicate "yes" by the three you would most like to have, and "no" by the three you least want.

The pamphlet titles were modeled after items from the PEC scale, six being liberal in direction, and six being conservative. Each pamphlet was associated with authorship such that a group expected to support Goldwater was said to author the conservative items and a group expected to support Johnson was said to author the liberal pamphlets. Examples are: (a) author—California Republican Council, title—"How Excessive Income Taxes Penalize Individual Success," and (b) author—Democratic National Committee, title—"Government's Responsibility for Providing Extra Income for the Needy." Although all the pamphlets and most of the author groups were fictional, they were apparently accepted as real. One subject who worked for a Republican group offered to supply the experimenter with free copies of the Republican pamphlets.

A few PEC items were slightly modified to bring them up to date. For example, "Most government controls over business should continue after the war" was shortened to read, "Most government controls over business should continue."

Procedure

The subjects were administered the questionnaire in groups of 10 or more from 2 weeks to 1 month prior to the 1964 Presidential election. They were told:

A survey is being conducted concerning various social attitudes and issues. Your frank responses to an opinion questionnaire will give an indication of how people feel about these issues. In the booklets that will be passed out in a moment, there are a number of questions for which we would like your opinions. Please be sure to respond to every question; in some cases you may not have a very strong opinion, but please make your best choice anyway for all questions.

Some parts of the questionnaire were collected before going on to later parts. This was done for all

² Parts a, c, and d of the questionnaire do not enter into the analyses reported in this paper. They were included to test a hypothesis concerning dissonance between voting for Goldwater and anti-Semitism. The data available were inadequate for making the test.

questionnaire parts except *e*, *f*, and *g* (PEC questions), which were all answered before being collected, and parts *i* (candidate's agreement or disagreement with themes) and *j* (pamphlets), which were turned in together at the end. This procedure was used to avoid the possibility of the subject changing his response to earlier parts due to knowledge of later parts. The total questionnaire took an hour or less to administer.

Dissonance between subject's conception of the candidates' conservatism and the conservatism attributed to the candidates was introduced by varying the themes of part *i* in which the candidates' alleged agreement or disagreement was given. In the zero-dissonance condition, Johnson was said to agree and Goldwater to disagree with the three liberal themes, and Goldwater was said to agree and Johnson to disagree with the three conservative themes. In the highest dissonance condition a total reversal occurred: Johnson was said to agree with the conservative themes and Goldwater to disagree, while Goldwater was said to agree with the liberal themes and Johnson to disagree. Dissonance levels other than zero and six derived from the remaining possible one to five reversals. The questionnaires were ordered such that the first one handed out contained zero reversals, the second contained one reversal, the third contained two reversals, etc., and zero reversals reappeared for every eighth questionnaire. A control group of 36 subjects received all of the questionnaire except parts *h* and *i*, the themes used to create dissonance.

RESULTS

The experiment was designed to make subjects perceive a contradiction between their conception of the candidates' political-economic-conservatism and assertions to the contrary. That this condition prevailed can be seen from two analyses. Scores on the PEC scale ranged from 1 on the liberal end to 7 on the conservative end. For Johnson voters, the average as Johnson would respond was 2.87, and as Goldwater would respond, 5.53 ($t = 21.80$, $df = 115$); for Goldwater voters, the means were 2.52 for Johnson and 5.57 for Goldwater ($t = 18.78$, $df = 44$). Responses to the set of six themes were also responses to PEC items. If the three conservative items were attributed to Goldwater and the three liberal items to Johnson, a score of 6 resulted. A complete reversal produced a score of 0. That Goldwater was considered conservative relative to Johnson is seen from the fact that Johnson voters obtained a mean of 5.24 and Goldwater voters, a mean of 5.23. There can be little doubt that

Goldwater was conceived as more conservative than Johnson.

As one would expect, Goldwater voters were more conservative than Johnson voters. Goldwater voters obtained a PEC mean of 4.52 when expressing their own views while Johnson voters obtained a mean of 3.99. This difference is significant beyond the .001 level ($t = 3.59$, $df = 159$).

Two criteria were established to ensure that the subjects were drawn from a population strongly committed to their candidate and to the belief that Goldwater is more conservative than Johnson. First, to ensure ego involvement in the issue, the analyses of pamphlet choices are based on only those subjects who said in part *b* of the questionnaire that they favored their choice strongly over his opponent. Second, to ensure that all scores in the analyses came from subjects who considered Goldwater more conservative than Johnson, no subject was included in the analyses unless his conception of Goldwater minus his conception of Johnson, as measured by PEC scores, was 1.5 or greater.

Some of the subjects selected on these criteria partially reversed themselves in their answers to the six PEC themes: out of 348 responses given by the 58 subjects meeting the criteria, 30 were either conservative items attributed to Johnson or liberal items attributed to Goldwater. To ensure a one-to-one correspondence between the 0-6 dissonance levels and the number of contradictions actually experienced, the dissonance category into which a given subject was placed was determined by taking the actual number of contradictions between the candidate to which the subject attributed the themes and the candidate to which the themes were attributed in the following section of the questionnaire. For example, a subject who attributed the three conservative themes to Goldwater, two of the liberal themes to Johnson, and one liberal theme to Goldwater still fell into the zero-dissonance group if the one reversed liberal theme also happened to be the only theme reversed in the next part of the questionnaire.

After experiencing the contradictions, it is expected that the subjects will seek pamphlets supporting their candidate and avoid

pamphlets supporting the opposition. The index of combined seeking and avoiding is the total score, which is distinguished from the seek score and the avoid score. Giving all consonant pamphlet choices a weight of +1 and all dissonant choices a weight of -1, a subject who seeks only consonant information and avoids only dissonant information will receive a total score of +6, and a subject who seeks only dissonant information and avoids only consonant information will receive a total score of -6. Other combinations of seeking or avoiding consonant or dissonant information will lead to intermediate total scores of 0 or even numbers ranging from -4 to +4. Individual seek scores and avoid scores were calculated in the same manner as total scores. Seek scores ranged from -3 (seeks only dissonant pamphlets) to +3 (seeks only consonant pamphlets). Avoid scores were determined in a similar way.

Because the dissonance level to which a subject was assigned depended upon his answers to the themes, the *N* per level could not be determined in advance. For dissonance levels 0, 1, and 5, the *N*s are small. Consequently, Level 0 was combined with Level 1 and Level 5 with 6 for the statistical analysis of mean differences. The means and *N*s by dissonance level are given in Table 1.

Two-factor analyses of variance were performed on total, seek, and avoid scores with dissonance level and candidate preference as

TABLE 1

MEANS FOR THE CONTROL GROUP AND GROUPS WITH DIFFERENT NUMBERS OF CONTRADICTIONS FOR TOTAL, AVOID, AND SEEK SCORES*

Score	No. contradictions					Control
	0,1	2	3	4	5,6	
Total						
<i>M</i>	-1.20	1.00	1.75	1.80	.15	2.58
<i>N</i>	10	12	8	10	13	24
<i>SD</i>	2.35	2.49	2.71	1.99	1.52	2.24
Avoid						
<i>M</i>	-.27	.17	1.50	.80	-.38	1.21
<i>N</i>	11	12	8	10	13	24
<i>SD</i>	1.85	1.59	1.77	1.16	1.26	1.74
Seek						
<i>M</i>	-.80	.43	.33	.60	.71	.50
<i>N</i>	10	14	9	10	14	24
<i>SD</i>	1.48	1.65	1.00	1.58	1.07	1.69

* *N*s differ among some corresponding score cells because a few subjects failed to indicate either titles most desired or least desired.

TABLE 2

ANALYSES OF VARIANCE FOR TOTAL, AVOID, AND SEEK SCORES WITH DISSONANCE LEVEL AND CANDIDATE PREFERENCE AS FACTORS

Factor	Total scores		Avoid scores		Seek scores	
	<i>SS</i>	<i>df</i>	<i>SS</i>	<i>df</i>	<i>SS</i>	<i>df</i>
Dissonance level	1	24.00**	1	6.46	1	14.00**
Candidate preference	4	1.16**	4	5.66*	4	2.00
Interaction	4	9.13	4	1.91	4	8.00***
Total	11	34.4	11	12.03	11	34.0

* Degrees of freedom differ for the three analyses because a few subjects failed to indicate either titles most desired or least desired.

** $p = .05$
*** $p = .01$

the factors. The analyses were least-squares solutions (Winer, 1962) to account for unequal *N*s. The results of these analyses are shown in Table 2. The analysis of total scores indicates a difference due to dissonance levels at less than the .05 level of significance. For avoid scores, dissonance levels differ at $p = .06$, but for seek scores differences between dissonance levels fail to reach $p = .10$. The difference due to candidate preference indicates that Goldwater voters exhibit greater seeking and avoiding than Johnson voters under the conditions of this experiment. The Goldwater voters may have been more committed to their views than the Johnson voters. The interaction for seek scores reflects a jagged line for Goldwater voters rising from below the line for Johnson voters at Level 0,1 and then remaining above the Johnson curve. Since there were only 16 Goldwater voters in the experimental group, there were only a few at each dissonance level. The small *N* makes the difference between jagged curves difficult to interpret.

Curves of dissonance versus total, seek, and avoid scores are shown in Figure 1. The curve of total scores approximates the inverted U predicted by Festinger. In understanding this outcome, it will be necessary to consider that the avoid curve is also an approximation of the expected shape, but the seek curve is not. The curves of total and avoid scores were subjected to trend analyses in which Level 0,1 is treated as .5 and 5,6 is treated as 5.5. The technique used (Gaito,

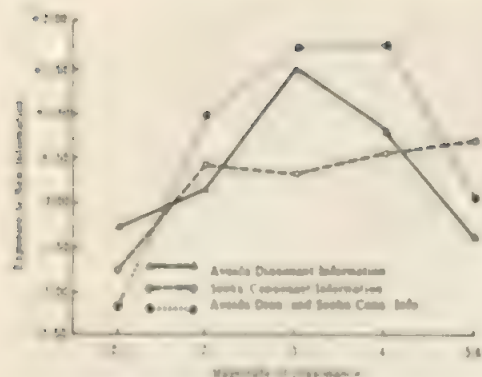


FIG. 1. Information seeking and avoiding as a function of dissonance.

1965) takes into account unequal N s and unequal intervals between points. A significant quadratic trend was found for total scores at almost the .001 level ($F = 12.49$, $df = 1,43$) and for avoid scores at less than the .01 level ($F = 7.67$, $df = 1,44$). Because it is not known whether the dissonance scale is more than an ordinal scale, these results are approximations. Consequently, the predicted differences between the peak and the lower and upper end points of the curve of total scores are of interest. These comparisons are significant by one-tailed t tests beyond the .005 and .05 levels. Similar tests for avoiding are significant beyond the .025 and .01 levels.

The mean of total scores for all subjects in the experimental sample is .64 against a control mean of 2.58, a difference significant beyond the .01 level ($t = 3.36$, $df = 75$). The difference between corresponding seek means of .30 and .50 is not significant at the .10 level although the higher mean is still for the control group. The difference between the avoid means of .26 for the experimental sample and 1.21 for the control group is significant beyond .05 ($t = 2.32$, $df = 76$). The curves tend to fall well below the empirical starting point indicated by the control group. The shape of the curve of total scores and the significant analyses of variance indicate that dissonance has an effect. But the comparison with the control group shows that the differential effect is lowered across dissonant levels.

DISCUSSION

The most general data total scores approximate an inverted U curve. Although the shape of that curve, together with the statistical analyses, supports the hypothesis, consideration of the control group and seek and avoid data requires further evaluation.

Why does the control group exhibit more information selectivity than the experimental sample? This question is of interest because a control group receiving the same treatment as the experimental subjects, except for the dissonance condition, is absent from previous dissonance research on seeking and avoiding. When total scores are split into seek and avoid scores, there is a difference between the experimental sample and the control group beyond the .05 level for avoid scores, but not for seek scores. The means for the experimental sample on avoid scores of .26 and on seek scores of .30 are quite similar. The corresponding control means are 1.21 and .50. Hence, the difference between the experimental sample and the control group appears due to a reduction in the avoidance behavior of the experimental sample relative to the control group. The only condition experienced by the experimental sample but not by the control group was the themes used to create dissonance. This experience apparently depressed the curve of the experimental sample.

Of the two sets of themes, the first set presents an added chance to rate the candidates on PEC items, which is similar to the experience both the control and experimental groups already had with the previous PEC items. The real difference between the experiences of the two groups comes from the second set of themes which authoritatively state the "actual" views of the candidates. Reduction of the overall level of avoidance would be explained by assuming that once an authoritative expression of the opposition's views is seen there is less reason for avoiding them. One can protect himself psychologically by avoiding the opposition's arguments, but after the person has been faced with these arguments, avoidance is a less effective psychological shield. In addition, some of the subjects may have decided after seeing the

themes that their initial impression of the candidates' relative conservatism was incorrect, and they may have reduced the need for avoidance by modifying their views. Since all the subjects had strong opinions about the candidates, they probably did not completely reverse themselves. It is also possible that the avoidance curve reflects curiosity motivation in addition to dissonance. Experiencing themes attributed to the candidates may make the subjects curious about the views of the nonpreferred candidate, thereby reducing avoidance behavior.

Another finding requiring explanation is the significant difference between dissonance levels for avoid scores but not seek scores. Festinger (1957), in stating the theoretical curves of seeking and avoiding, specified only the general nature of the relationship. He specified that both curves were in the general shape of an inverted U with an initial rise being more gradual than the final drop, and with the seek curve reaching a higher peak than the avoid curve. He also predicted that the seek curve would decrease to an indifference point at the limit of dissonance and the avoid curve would fall to an indifference point near the limit of dissonance and then rise as subjects switch from avoiding to seeking. It is probably very difficult to drive subjects to the limit of dissonance in a laboratory setting where a powerful, real issue is used. In any case, in this experiment it is not known if the limit of dissonance was reached, nor is the location of the indifference point determined. Therefore, the discussion is

restricted to the general inverted U shape of the two curves.

The data of this experiment are used to suggest further specifications of the nature of the predicted curves. Figure 2 shows the general relationship between dissonance and seek and avoid responses. The dashed line is the seek curve, and the solid line is the avoid curve. The dotted tail represents the prediction that near the limit of dissonance a subject will switch from avoiding to seek dissonant information. In order to fit the data to the theoretical curves, the avoid curve is made more peaked than the seek curve. Although this is different from Festinger's original formulation, it is consistent with findings in conflict studies (Miller, 1944) where the gradient of avoidance is steeper than the gradient of approach.

Two scales are shown on the abscissa of Figure 2. One is a hypothetical scale of "true" units of dissonance. Its precise nature is unknown. Three things can be said about it: (a) Both seek and avoid curves rise from the base line at the true zero point (indifference point); (b) the seek curve returns to the base line near the limits of dissonance, and avoidance decreases to the base line at a high dissonance level not quite so close to the upper limit; (c) the parts of the curves studied in this experiment will probably be only a portion of the total curve, and that portion will probably fall somewhere between the upper and lower limits of dissonance. In Figure 2, the portion of the curves between the vertical lines represents the portion of the theoretical curves assumed to underlie the findings of this experiment. The second scale on the abscissa is the 0.6 levels of dissonance used in this experiment. Zero does not mean no dissonance. It means no added dissonance due to contradictions between sets of themes.

The difference among avoidance scores due to dissonance would show up even with the crude measures available if the peaked portion of the avoidance curve was tapped. The two tails of the portion of the avoid curve between the vertical lines are different enough from the peak to determine a low p value from an analysis of variance. The data plotted in Figure 1 do show the avoid curve to

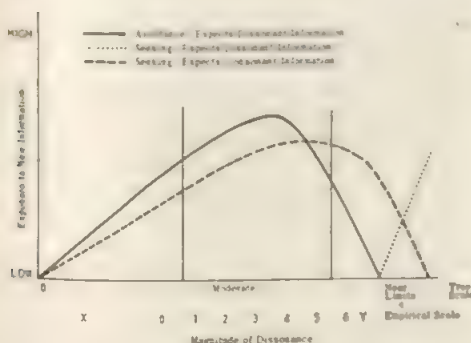


FIG. 2. Hypothetical curves of seeking and avoiding as a function of dissonance.

be considerably more peaked than the seek curve. The rise in the portion of the seek curve between the vertical lines of Figure 2 is quite gradual, and it would take a sensitive measure before this rise would be reliably detected. The seek curve in Figure 1 does tend to rise quite gradually, and this rise does not produce a statistically significant result.

The data of this experiment differ from what is more commonly found in that a significant difference due to avoiding dissonant information occurs, but not due to seeking consonant information. Brehm and Cohen (1962), after reviewing several studies, concluded that the reverse is usually found, and Brock (1965) felt this finding was compelling enough to require a refinement of dissonance theory. Examination of Figure 2 provides a possible explanation for the finding of an avoidance effect. Previous studies where consonant seeking is found have used at most two levels of dissonance. Suppose points like X and Y on the empirical dissonance scale are tapped. A difference would be predicted in seeking consonant information since seeking at Point X is considerably less than at Point Y. But no difference is expected in avoiding dissonant information because avoiding at Point X is virtually the same as at Point Y. If dissonance levels 0, 1, and 2 are combined into a single low-dissonance group, and levels 4, 5, and 6 into a high-dissonance group, the more usual finding obtains. Consonant information is more often sought by the high group ($t = 2.21$, $df = 46$), but there is no significant difference in avoiding dissonant information ($t < 1$).

Although the foregoing interpretation accounts for the results, other alternatives exist. First, we cannot be sure that pamphlet choice is an index of both seeking and avoiding. It is possible that the seek curve is relatively flat because the method employed tapped avoidance and not seeking. Second, we cannot be certain that the data reflect results from the middle range of the dissonance scale. The explanation based on Figure 2 would not hold if only the lower or upper range of dissonance was sampled. Third, the seek curve could be due to a mixture of dissonance and curiosity motivation. Suppose seeking due to curiosity is a positively sloped,

linear function of the number of contradictions. A linear curve added to an inverted U could approximate the gradually rising seek data. Fourth, the upper end of the avoid curve might arise because subjects do not believe the five or six contractions. If they do not believe, dissonance due to the contradictions would not be expected. Then the subjects receiving five or six contradictions should exhibit the same degree of avoidance as those receiving no contradictions.

REFERENCES

- ADAMS, J. S. Reduction of cognitive dissonance by seeking consonant information. *Journal of Abnormal and Social Psychology*, 1961, **62**, 74-78.
- ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, N. R. *The authoritarian personality*. New York: Harper, 1950.
- BREHM, J. W., & COHEN, A. R. *Explorations in cognitive dissonance*. New York: Wiley, 1962.
- BROCK, T. C. Commitment to exposure as a determinant of information receptivity. *Journal of Personality and Social Psychology*, 1965, **2**, 10-19.
- BRODBECK, N. The role of small groups in mediating the effects of propaganda. *Journal of Abnormal and Social Psychology*, 1956, **52**, 166-170.
- CANON, L. K. Self-confidence and selective exposure to information. In L. Festinger (Ed.), *Conflict, decision, and dissonance*. Stanford: Stanford University Press, 1964. Pp. 83-95.
- CHAPANIS, N. P., & CHAPANIS, A. Cognitive dissonance: Five years later. *Psychological Bulletin*, 1964, **61**, 1-22.
- COHEN, A. R., BREHM, J. W., & LATANÉ, B. Change of strategy and involuntary exposure to information under public and private conditions. *Journal of Personality*, 1959, **27**, 63-73.
- EHRlich, D., GUTTMAN, I., SCHÖNBACH, P., & MILLS, J. Postdecision exposure to relevant information. *Journal of Abnormal and Social Psychology*, 1957, **54**, 98-102.
- FEATHER, N. T. Cigarette smoking and lung cancer: A study of cognitive dissonance. *Australian Journal of Psychology*, 1962, **14**, 55-64.
- FEATHER, N. T. Cognitive dissonance, sensitivity, and evaluation. *Journal of Abnormal and Social Psychology*, 1963, **66**, 157-163.
- FESTINGER, L. *A theory of cognitive dissonance*. Stanford: Stanford University Press, 1957.
- FESTINGER, L. *Conflict, decision, and dissonance*. Stanford: Stanford University Press, 1964.
- FREEDMAN, J. L. Confidence, utility, and selective exposure to information: A partial replication. *Journal of Personality and Social Psychology*, 1965, **2**, 778-780. (a)
- FREEDMAN, J. L. Preference for dissonant information. *Journal of Personality and Social Psychology*, 1965, **2**, 287-289. (b)

- FREEDMAN, J. L., & SEARS, D. O. Voters' preferences among types of information. *American Psychologist*, 1963, 18, 375. (Abstract)
- FREEDMAN, J. L., & SEARS, D. O. Selective exposure. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. New York: Academic Press, 1965. Pp. 57-97.
- GAITO, J. Unequal intervals and unequal n in trend analysis. *Psychological Bulletin*, 1965, 65, 125-127.
- JECKER, J. D. Selective exposure to new information. In L. Festinger (Ed.), *Conflict, decision, and dissonance*. Stanford: Stanford University Press, 1964. Pp. 65-81.
- MACCOBY, E. E., MACCOBY, N., ROMNEY, A. K., & ADAMS, J. S. Social reinforcement in attitude change. *Journal of Abnormal and Social Psychology*, 1961, 63, 109-115.
- MILLER, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. 1. New York: Ronald Press, 1944. Pp. 431-465.
- MILLS, J. Avoidance of dissonant information. *Journal of Personality and Social Psychology*, 1965, 2, 589-593. (a)
- MILLS, J. Effect of certainty about a decision upon postdecision exposure to consonant and dissonant information. *Journal of Personality and Social Psychology*, 1965, 2, 749-752. (b)
- MILLS, J. The effect of certainty on exposure to information prior to commitment. *Journal of Experimental Social Psychology*, 1965, 1, 348-355. (c)
- MILLS, J., ARONSON, E., & ROBINSON, H. Selectivity in exposure to information. *Journal of Abnormal and Social Psychology*, 1959, 59, 250-253.
- MILLS, J., & ROSS, A. Effects of commitment and certainty upon interest in supporting information. *Journal of Abnormal and Social Psychology*, 1964, 68, 552-555.
- ROSEN, S. Postdecision affinity for incompatible information. *Journal of Abnormal and Social Psychology*, 1961, 63, 188-190.
- SEARS, D. O. Biased indoctrination and selectivity of exposure to new information. *Sociometry*, 1965, 4, 363-376.
- SEARS, D. O. Opinion formation and information preferences in an adversary situation. *Journal of Experimental Social Psychology*, 1966, 2, 130-142.
- SEARS, D. O., & FREEDMAN, J. L. Commitment, information utility, and selective exposure. United States Navy Technical Report No. 12, 1963, Office of Naval Research, Nonr-233 (54) NR 171-350.
- SEARS, D. O., & FREEDMAN, J. L. Effects of expected familiarity with arguments upon opinion change and selective exposure. *Journal of Personality and Social Psychology*, 1965, 2, 420-426.
- STEINER, I. D. Receptivity to supportive versus non-supportive communications. *Journal of Abnormal and Social Psychology*, 1962, 65, 266-267.
- WINER, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962. Pp. 291-297.

(Received June 16, 1966)

OBSERVATION AND REHEARSAL AS DETERMINANTS OF PROSOCIAL BEHAVIOR¹

DAVID ROSENHAN

AND

GLENN M. WHITE

Center for Psychological Studies

Princeton University

Educational Testing Service, Princeton, New Jersey

In a study in altruistic behavior, 4th- and 5th-grade Ss played a bowling game, once in the presence of an adult model and once in his absence. Each time the model won gift certificates, he donated $\frac{1}{2}$ of them to a charity. No control Ss, who did not observe a model, contributed to the charity while playing alone. Among Ss who observed the model, it was primarily those Ss who contributed in the model's presence who also contributed in his absence, suggesting that rehearsal as well as observation were necessary for the elicitation of this phenomenon. The valence (positive or negative) and occurrence of a prior relationship with the model had peculiar and perhaps indeterminate effects on the elicitation of altruistic behavior.

What are the conditions that lead people to evidence concern for the welfare of others: parents, say, for children; the healthy for the sick; civil rights workers for Negroes? These acts reflect altruism which, in the original meaning that Auguste Comte gave to the word, implies concern for others. Altruism has not received much attention in the psychological literature.

In part, the nature of altruism may have escaped attention because psychological formulations of personality strongly imply that men have no concern outside of themselves. Theories of personality almost invariably conceive of it as "something integumented, as residing within the skin [Allport, 1960, p. 306]." Consequently, altruistic behavior has been conceptualized by psychologists as either selfish or neurotic in origin. Anna Freud (1937), for example, considered that altruism arises from deprivation and inhibition, from reaction formations to aggression, or from the expiatory dynamics of guilt. Glover (1925) viewed it as deriving from oral character traits, while Fenichel (1945) related it to homosexuality and castration anxiety. Learning theories, with their strong emphasis

on egocentric tension, need reduction, or drive reduction, also have obvious limitations for conceptualizing phenomena that are apparently so directed to reducing others' needs.

This egocentric view of personality is gradually changing, and the fact that altruistic attitudes and behavior are becoming matters of more substantial concern is seen in more recent writings (cf. Berkowitz & Daniels, 1963, 1964; Fromm, 1947; Lewin, 1951; Peck & Havighurst, 1960). Sociologists who in general have been more comfortable with open-system dynamics have conceived of norms of reciprocity (Gouldner, 1960; Homans, 1961; Sorokin, 1950) by which to explain the willingness to be helpful to others.

The present experiment explores the antecedents of charitable behavior. Beginning from the premise that altruistic behaviors are derived from internalized norms and are conceptually linked by the individual's willingness to give up more than he gains, the authors hypothesized that the norm is acquired and/or elicited on the basis of observation. The role of observation is clearly seen in the acquisition and elicitation of such behaviors as aggression (Bandura, 1965), internalized standards of self-reward (Bandura, 1965; Bandura & Whalen, 1966; Mischel & Liebert, 1964; Rosenhan & Frederick, 1966), aversive behaviors (Mischel & Grusec, 1966), and moral judgments (Bandura, 1965). Its direct relevance to altruistic behavior is suggested by two studies. First, in a retrospective

¹ This research was supported in part by Grant 1-P01-HD 01762 from the National Institute of Child Health and Human Development. We are grateful to Anne Burrowes, Irene Kostin, and Terrence Tankersley who assisted in the pilot work for, and served as observers during, this experiment. Jerome E. Singer and Richard Peterson commented on an earlier draft of this report.

study Tomkins (1965) observed that the leading advocates of the abolition of slavery had themselves observed and been influenced by other altruists. Moreover, a recent *in vivo* study of civil rights workers (Rosenhan, 1966) indicated that individuals who remained committed to and active in civil rights work over an extended period reported close affective ties to at least one parent who was altruistic. The present study may be seen as an attempt to examine these findings under controlled laboratory conditions.

The extent to which altruistic behavior is evidenced following observation of an altruistic model should be influenced by the nature of the interaction of model and child. The direction of the influence, however, is not clear. If internalization of moral norms has its roots in the defensive posture, a commonly held position (cf. Freud, 1937) that receives some support in the work of Aronfreed (1964) and Aronfreed and Reber (1965), then interaction with a punitive model should facilitate altruistic behavior. On the other hand, there is strong evidence indicating that a prior nurturant relationship facilitates internalization (Allinsmith, 1960; Bandura, 1965; Maccoby, 1959; Sears, Maccoby, & Levin, 1957; Whiting, 1959). The further observation that parents of altruistic children typically are warm, nurturant, and relatively nonpunitive (Peck & Havighurst, 1960) leads to the prediction that a prior positive relationship would be superior to a negative one in eliciting altruistic behavior. In the present experiment the model was either critical or supportive of the subject during a brief interaction, and the effects of these relationships, and of no prior relationship, were examined in the subject's subsequent behavior.

METHOD

Subjects and Model

Sixty-five boys and 65 girls drawn from the fourth and fifth grades of two middle-class public schools served as subjects.² Two additional subjects were eliminated when the apparatus failed during

their session. The subjects were sent to the experimental room in boy-girl order as determined by their teacher on the basis of availability and were assigned to the four preexperimental conditions. A total of 20 boys and 20 girls were assigned to each of these conditions: negative reinforcement (NR), positive reinforcement (PR), no interaction (NI). In addition, 5 boys and 5 girls were assigned to a no-model control condition (NM).

The model was a 25-year-old male graduate student.

Materials and Apparatus

Concern on the part of school officials precluded the use of money as the commodity with which the subjects could be altruistic; therefore, a money surrogate was employed which had previously been demonstrated to be highly valued by fourth and fifth graders. Gift certificates of \$.05 redeemable at a popular and nearby toy store on parental signature were employed.³

Performance in a miniature bowling game similar to the one described by Bandura and Whalen (1966) served as the task through which the subject could win the gift certificates. The game was played by rolling a marble along a 3-foot ramp, the end of which was shielded from the subject's view. As the marble rolled off the end of the ramp, a switch was triggered which activated one of four lights located behind the plastic numerals 5, 10, 15, and 20. The numerals were embedded in a vertical sheet of wood which faced the subject at the end of the ramp. A game consisted of 20 rolls and was automatically terminated by a buzzer. The sequence in which the numerals were illuminated was predetermined and remained constant for all games.

Ten piles of two \$.05 gift certificates were placed on a table near the bowling game. Also on the table was an open box with a sign which read, "Trenton Orphans Fund," and displayed pictures of children in ragged attire. Ten certificates were maintained in the box, both to serve as a cue for giving and to minimize the possibility that the subject would be influenced to give by the expectation that the experimenter or the model would notice later whether or not certificates were in the box.

Procedure

As the subject entered the experimental room, a female experimenter obtained the subject's name and introduced the model to the subject as the inventor of the game which the subject would later be allowed to play. All subjects were allowed to play, both in the presence and absence of the model. At no time was the experimenter present during the games.

Negative reinforcement (NR). After having introduced the model to the subject, the experimenter stated,

³ We are grateful to Neisner's Department Store, Trenton, New Jersey, for their cooperation in processing the gift certificates.

² We are grateful to the administration and teachers of the Lawrence Township school system and particularly to Irvin Hyman, school psychologist, and to Esther Updike and Neal Young, principals, for their cooperation.

I have to go and get a few more gift certificates, then we'll be ready to play. Why don't you two just sit down for a minute and get acquainted? I'll be right back, and then you can play the bowling game.

As the experimenter left, the model engaged the subject in a conversation by asking questions and making responses contingent upon the subject's answers. During a 5-minute period, the model made nine negatively reinforcing statements, maintaining a scowling facial expression and a curt verbal style throughout the interaction period. Typical negatively reinforcing statements were:

What grade are you in? [Answer] What's the matter? Were you held back a year? [also] Do you have any hobbies? [If the subject said he had no hobby] Well everybody should have a hobby, why don't you? [If the subject offered one] I don't think hobbies like that are very useful; why don't you have a constructive hobby, like _____?

Although each subject in this condition was negatively reinforced nine times, the interview was kept somewhat flexible, allowing the subject to elaborate when he wished to do so. The sequence of questions and statements was altered when doing so promised to make the interaction seem more natural. Subjects characteristically turned away from the model during this condition, suggesting some discomfort with him. There was no evidence, however, of any distress that the subjects could not accommodate.

After 5 minutes had elapsed, the experimenter returned with the gift certificates and administered the game instructions.

Positive reinforcement. (PR). The same procedure was followed in this condition as was followed in the NR condition. The questions and statements employed were parallel to the ones used in NR, with the positive reinforcement supplanting the negative reinforcement. Examples parallel to the ones cited above are:

What grade are you in? [Answer] Have you skipped a grade? [also] Do you have any hobbies? [If the answer was no] Very good! I'll bet you have too many important things to do to spend time with hobbies. [If the subject offered one] Very good! I think that is an excellent hobby.

The model smiled frequently at the subject throughout the interview and was generally approving and permissive.

No-interaction control (NI). This condition was included to control for the possibility that prior interaction with a model of any kind might affect giving. In this condition, as well as in NR and PR, subjects were provided with an altruistic model. The only distinction between this condition and the above two conditions was the exclusion of the 5-minute interaction period in NI.

No-model control (NM). Ten subjects were assigned to this condition in order to determine whether subjects would give in the absence of a model and on the basis of situational cues and the experimenter's instructions alone. NM subjects were not provided with a model, although a model observed while the subject played the first game, allowing the subject to give in the model's presence.

Instructions

All bowling-game and charity-box instructions were administered by the experimenter. After the mechanics of the game were explained, it was made clear to the subject that obtaining a "20" by rolling the marble constituted winning. The prize for each 20 was two \$.05 gift certificates.

Except in the NM condition, the subject was told that the model would play the first game with the subject in order to make sure that the bowling game was working properly. NM subjects were told that the model would *watch* the subject play the first game to see that the machine was in proper working order. The experimenter informed all subjects that they would be allowed to play a second game *alone* and that they were to go back to class after finishing that game.

After the experimenter ascertained that the subject understood how to play the game, she casually, as if as an afterthought, called the subject's and the model's attention to the charity box, explaining that some gift certificates were being collected for some orphans and that if "... either of you would like to give some of your certificates to them you can, but you do not have to."

Model-Present Game

In all conditions except NM, the model and the subject alternated turns, with the model winning on Trials 1 and 5, and the subject on Trials 12 and 18. A game consisted of 20 rolls, 10 each for the model and the subject. Throughout the game, the model feigned intense interest in the game, paying minimal attention to the subject. The model's only verbalizations consisted of commenting that in order to win, one must hit the right spots at the proper speeds, repeating the numbers as they flashed on, and after winning the first time saying (ostensibly to himself), "I won. I believe I will give one certificate to the orphans each time I win." Having said this, the model picked up a stack of certificates, dropped one into the charity box, and kept the second. Upon winning the second time, the model again picked up two certificates and dropped one into the box, saying nothing this time. Because the subject was standing between the model and the charity box, the model had to reach in front of the subject to donate, assuring thereby that the subject observed the model giving.

When the subject won, the model was either in the process of examining the marble return mechanism or intensely gazing at the illuminated numerals, thus avoiding looking at the subject and inadvert-

providing the subject with cues as to whether or not he should give. The model also avoided seeing whether or not the subject gave, minimizing the possibility that social reinforcement accompanied either giving or not giving. The charity box was situated such that the model could notice whether or not the subject gave in the model's presence without enabling the subject to see that he was being observed. At the conclusion of the first game, the model left the subject alone, commenting as he left that the subject had one game left to play but that he, the model, had to leave and would not be back, and for the subject to go back to class when he had finished. The model closed the door behind him as he exited. The subject then played for 20 trials alone. After the subject left, it was determined whether or not he had given in the model's absence by counting the certificates in the charity box.

RESULTS

The results supported the hypothesis that exposure to a giving model would elicit substantially more altruistic behavior than would occur under similar circumstances with no model. Of the 120 subjects who observed the model give, 57 (47.5%) were altruistic, that is, gave in the model's absence. On the other hand, of the 10 NM subjects, none evidenced altruistic behavior. A comparison of these two proportions employing Fisher's exact-probability test (S. Siegel, 1956) yielded a $p = .0046$ (two-tailed) as seen in Table 1.

The 120 subjects who were exposed to the model were categorized according to whether or not they gave in the model's presence (conformity) and/or his absence (internalization). Table 2 reveals that 63% of the subjects conformed and gave in the model's presence, while 48% internalized, giving in the model's absence. Of the latter subjects, nearly 90% had previously given in the

TABLE 1

ALTRUISTIC RESPONSES OF SUBJECTS WHO WERE OR WERE NOT EXPOSED TO MODELS

Gave in model's absence?	Exposed to model		
	Yes	No	Totals
Yes	57	0	57
No	63	10	73
Totals	120	10	130

model's presence. A chi-square analysis employing Yates' correction for continuity yielded $\chi^2 = 29.84$, $p < .001$. McNemar's test for significance of change (S. Siegel, 1956) indicated that significantly more subjects changed from givers in the model's presence to nongivers in his absence than vice versa ($\chi^2 = 10.45$, $p < .005$, corrected for continuity).

Having found that observing a model promotes internalization of prosocial behavior, we now turn to the question of whether the prior relationship with the model has differential effects, with regard to both conformity in the presence of the model and internalization in his absence. Overall, the findings were quite disappointing. Chi-square analyses failed to reveal differential effects, in both the model-present and model-absent conditions, for positive, negative, or no relationship. The authors sensed in examining these data, however, that there were some suggestive and potentially interesting findings embedded in the tables, and undertook a partitioned chi-square analysis (Castellan, 1965) on the matrices described in Figure 1.

While there appeared to be no differences for either boys or girls in the effects of positive versus negative prior relationships, at-

TABLE 2

RELATIONSHIP BETWEEN GIVING IN THE PRESENCE AND THE ABSENCE OF A MODEL

Gave in model's absence? (internalization)	Gave in model's presence? (conformity)						All Ss
	Yes			No			
	Boys	Girls	Total	Boys	Girls	Total	
Yes	28	23	51	3	3	6	57
No	9	16	25	20	18	38	63
Total	37	39	76	23	21	44	120 ^c

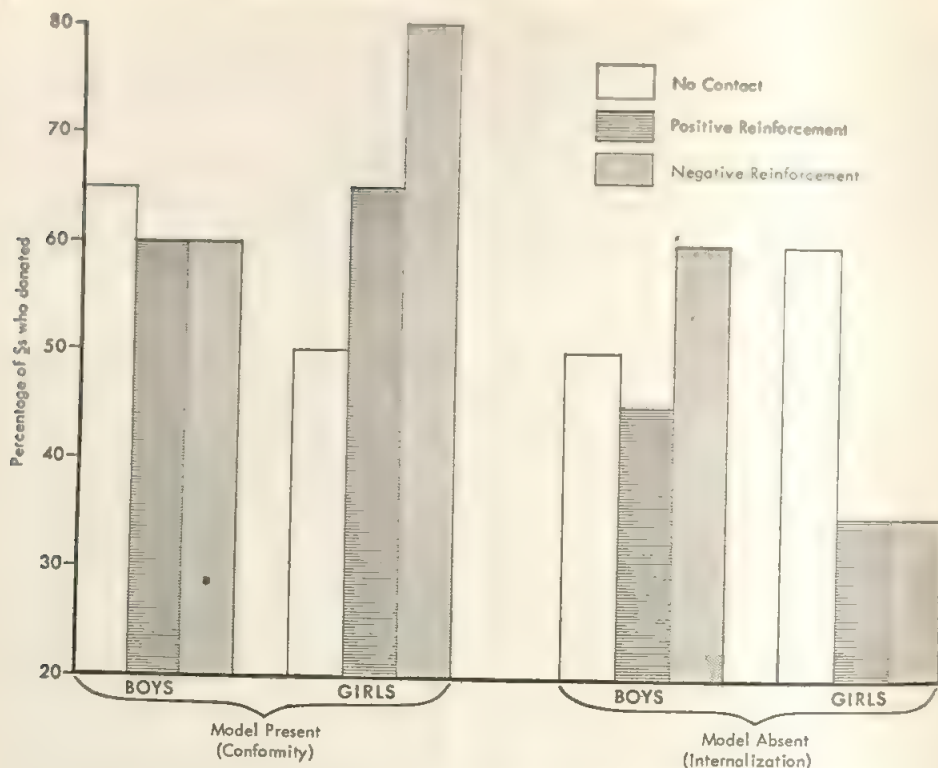


FIG. 1. Conformity and internalization as a function of prior relationship with the model.

tention was drawn to a comparison of those who had had a relationship with the model (regardless of valence) versus those who had not. Note that the presence or absence of the model seemed to moderate the effects of prior relationship among girls but not among boys. When the model was present, more girls who had had a prior relationship with him donated than those who had not ($\chi^2 = 2.97, .10 < p > .05$). However, when the model was absent the findings were reversed: more girls who had *not* had a prior relationship gave than those who had ($\chi^2 = 3.39, .10 < p > .05$). The effects of prior relationship held up equally well when the behavior of girls was compared to boys. When the model was present, more girls (72%) contributed than boys (60%). In the absence of the model, however, fewer girls (35%) than boys (52%) contributed. To the extent that the reader is struck by the special quality of these relationships, it must be borne in mind that the model was male, and that the results might have been different had we employed

a female model. At best, as we noted above, the obtained relationships are merely suggestive.

That the gift certificates were highly valued by the subjects was evidenced by the fact that 16 of the 130 subjects stole one or more certificates, usually from extra certificates on the table, but in some instances also from the charity box. A hidden observer who watched and kept notes on the behavior of 65 subjects recorded that many of the subjects engaged in considerable ambivalent behavior, indicative of approach-avoidance conflict in the model-absent condition, when deciding whether to give.

DISCUSSION

It should be emphasized that what is referred to as altruism is the voluntary dispensing of a valued commodity to a needy other without the clear prospect of being subsequently reinforced. That nearly 50% of the subjects donated their gift certificates to

charity in the absence of the model indicates that altruistic behavior *can* be fairly common and that it can be elicited in the laboratory.

The model made it quite clear that there would be no further contact between himself and the subject, and reinforced this notion in a variety of salient ways. That the subjects donated despite this abrupt termination of relationship suggests that there are conditions under which subjects will be charitable even if the prospect of subsequent reinforcement is minimal or nonexistent. It also seemed clear from these findings that the perception of dependency *alone* did not result in altruistic behavior (Berkowitz & Daniels, 1963, 1964). None of the NM control subjects donated, even though they had been instructed that they could donate if they so desired and had observed the same charity box and pictures of ragged children as had the other subjects (Table 1).

Observation and Rehearsal

The permissive instructions that were used in this experiment are similar to those that are often employed in the teaching of prosocial behavior in educational environments. By themselves, they were unsuccessful for internalizing a prosocial norm. And while it might be useful to explore the effects of other kinds of instructions (less permissive ones, for example), it is clear that observation of a model is one powerful determinant of altruistic behavior.

Powerful as observation of a model is, it is not sufficient, since many subjects failed to donate in the model's absence. There is excellent reason to believe that these nondonors *saw* the model give, both because the model had to reach across the subject to get to the charity box and because many nondonors exhibited considerable hesitancy and ambivalence when, while alone, they won but failed to donate. Since it was primarily those subjects who donated in the model's presence who also donated in his absence, it was concluded that observation *and* rehearsal may be necessary for the internalization of the altruistic norm.

If this reasoning is correct, then the dynamics of acquisition and elicitation of altruistic behavior differ from, for example, those

of aggressive behavior where merely observing a model is sufficient for acquisition and elicitation (Bandura, 1965; Lovaas, 1961; A. Siegel, 1956). Two hypotheses may reasonably account for the difference. First, there are rewards for vigorous aggression, particularly those of strong proprioceptive feedback and possibly others that exist at the fantasy level (hero, power). These rewards find no counterpart in altruistic behavior where giving up something, particularly something that is prized, is presumably distasteful and negatively reinforcing. Rehearsal in the presence of the model may assist in establishing the *habit* of altruism, perhaps because the subject assumes that the model approves of his behavior, and because such construed approval outweighs and compensates for the material losses.

A second hypothesis would examine the degree to which the disposition to be altruistic has been learned relative to other dispositions. Consider aggression again. While novel aggressive responses can be acquired on the basis of observation, one can argue that the disposition to be aggressive has been both overlearned and strongly inhibited in our culture. Children constantly observe aggression "live"—in films, on television—and they have presumably learned something of its rewards and consequences. Therefore, observation alone, under permissive conditions, is sufficient to elicit aggression. Altruistic behavior, however, is neither so evident nor so overlearned in our culture. Its rewards are neither immediate nor apparent, and on those occasions when children do witness altruism, it sometimes elicits a negative or cynical remark from other observers.

Prior Interaction with the Model

In general, there were no visible effects of positive and negative relationships with the model on subsequent internalization of altruistic norms. While it may in fact be the case that the quality of the prior relationship has little bearing on internalization, it may also be true that the experimental treatment was too brief and innocuous to elicit the potential effects.

There was suggestive evidence in this experiment that prior interaction with the

model affected the behavior of girls but not of boys. For the former, prior interaction was associated with more donating in the model's presence, but less in his absence. This dependence upon external sanctions among middle-class girls and the relative independence from external events among boys has also been observed by Aronfreed (1961) with regard to moral responses to transgression, by Terman (1925) with regard to honesty behaviors, and by Hartshorne and May (1928), who found that girls were less resistant to temptation than were boys. That girls have the superior reputation in our society for moral sensitivity in general, and for kindness, generosity, and charity in particular, may, as Terman and Tyler (1954) suggested, reflect more on girls' sensitivity to these norms and on their verbal facility regarding them than on the degree to which they have internalized them. One must recall, however, that these data were obtained with a male model, and that it is not yet appropriate to extrapolate these findings to girls who observe a female model.

REFERENCES

- ALLINSMITH, W. The learning of moral standards. In D. R. Miller & G. E. Swanson (Eds.), *Inner conflict and defense*. New York: Holt, 1960. Pp. 141-176.
- ALLPORT, G. W. The open system in personality theory. *Journal of Abnormal and Social Psychology*, 1960, **61**, 301-310.
- ARONFREED, J. The nature, variety, and social patterning of moral responses to transgression. *Journal of Abnormal and Social Psychology*, 1961, **63**, 223-240.
- ARONFREED, J. The origin of self-criticism. *Psychological Review*, 1964, **71**, 193-218.
- ARONFREED, J., & REBER, A. Internalized behavioral suppression and the timing of social punishment. *Journal of Personality and Social Psychology*, 1965, **1**, 3-16.
- BANDURA, A. Vicarious processes: A case of no trial learning. In L. Berkowitz (Ed.), *Advances in social psychology*. Vol. 2. New York: Academic Press, 1965. Pp. 1-55.
- BANDURA, A., & WHALEN, C. K. The influence of antecedent reinforcement and divergent modeling cues on patterns of self-reward. *Journal of Personality and Social Psychology*, 1966, **3**, 373-382.
- BERKOWITZ, L., & DANIELS, L. R. Responsibility and dependency. *Journal of Abnormal and Social Psychology*, 1963, **66**, 429-436.
- BERKOWITZ, L., & DANIELS, L. R. Affecting the salience of the social responsibility norm: Effects of past help on the response to dependency relationships. *Journal of Abnormal and Social Psychology*, 1964, **68**, 275-281.
- CASTELLAN, J. N., JR. On the partitioning of contingency tables. *Psychological Bulletin*, 1965, **64**, 330-338.
- FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.
- FREUD, A. *The ego and the mechanisms of defense*. London: Hogarth, 1937.
- FROMM, E. *Man for himself: An inquiry into the psychology of ethics*. New York: Holt, 1947.
- GLOVER, E. Notes on oral character formation. *International Journal of Psycho-analysis*, 1925, **6**, 131-154.
- GOULDNER, A. W. The norm of reciprocity: A preliminary statement. *American Sociological Review*, 1960, **25**, 161-178.
- HARTSHORNE, H., & MAY, M. A. *Studies in the nature of character*. Vol. 1. *Studies in deceit*. New York: Macmillan, 1928.
- HOMANS, G. C. *Social behaviour: Its elementary forms*. New York: Harcourt, Brace, 1961.
- LEWIN, K. *Field theory in social science*. New York: Harper, 1951.
- LOVAAS, I. J. Effect of exposure to symbolic aggression on aggressive behavior. *Child Development*, 1961, **32**, 37-44.
- MACCOBY, E. E. The generality of moral behavior. *American Psychologist*, 1959, **14**, 358.
- MISCHEL, W., & GRUSEC, J. Determinants of the rehearsal and transmission of neutral and aversive behaviors. *Journal of Personality and Social Psychology*, 1966, **3**, 197-205.
- MISCHEL, W., & LIEBERT, R. M. Effects of discrepancies between observed and imposed reward criteria on their acquisition and transmission. *Journal of Personality and Social Psychology*, 1964, **3**, 45-53.
- PECK, R. F., & HAVIGHURST, R. J. *The psychology of character development*. New York: Wiley, 1960.
- ROSENHAN, D. *The origins of altruistic social autonomy*. Princeton, N. J.: Educational Testing Service, 1967, in press.
- ROSENHAN, D., FREDERICK, F., & BURROWES, A. *Preaching and practicing: The consequences of consistencies and discrepancies in imposed and observed standards*. (Res. Mem. 66-27) Princeton, N. J.: Educational Testing Service, 1966.
- SEARS, R. R., MACCOBY, E. E., & LEVIN, H. *Patterns of child rearing*. Evanston, Ill.: Row, Peterson, 1957.
- SIEGEL, A. E. Film-mediated fantasy aggression and strength of aggressive drive. *Child Development*, 1956, **27**, 365-378.
- SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.
- SOROKIN, P. *Altruistic love*. Boston: Beacon Press, 1950.
- TERMAN, L. M. *Genetic studies of genius*. Vol. 1. *Mental and physical traits of a thousand gifted*

- children. Stanford: Stanford University Press, 1925.
- TIEMAN, L. M., & TYLER, L. E. Psychological sex differences. In L. Carmichael (Ed.), *Manual of child psychology*. (2nd ed.) New York: Wiley, 1954. Pp. 1104-1114.
- TOMKINS, S. S. The psychology of commitment: I. The constructive role of violence and suffering for the individual and for his society. In S. S. Tomkins & C. E. Izard (Eds.), *Affect, cognition, and personality*. New York: Springer, 1965.
- WHITING, J. W. M. Sorcery, sin, and superego: A cross cultural study of some mechanisms of social control. In M. R. Jones (Ed.), *Nebraska symposium on motivation, 1959*. Vol. 7. Lincoln: University of Nebraska Press, 1959. Pp. 174-195.

(Received June 27, 1966)

Manuscripts Accepted for Publication in the
Journal of Personality and Social Psychology

- Conceptual Complexity and Remote Associative Proficiency as Creativity Variables in a Complex Problem-Solving Task: Marvin Karlins*: Department of Psychology, Princeton University, Princeton, New Jersey 08540.
- Dogmatism, Prenegotiation Experience, and Simulated Group Representation as Determinants of Dyadic Behavior in a Bargaining Situation: Daniel Druckman*: Institute for Juvenile Research, 608 South Dearborn, Chicago, Illinois 60605.
- Personality Determinants of Psychophysiological Response to Stress: A Theoretical Analysis and an Experiment: Edward M. Opton, Jr., Elizabeth Alfert, and Richard S. Lazarus*: Department of Psychology, University of California, Berkeley, California 94720.
- Personality, Defensiveness, and Educational Achievement: Binyamin Z. Lebovits* and Adrian M. Ostfeld: University of Illinois, Medical Center, 1853 West Polk Street, P. O. Box 6998, Chicago, Illinois 60680.
- Valence in the Adoption of Solutions by Problem-Solving Groups: II. Quality and Acceptance as Goals of Leaders and Members: L. Richard Hoffman* and Norman R. F. Maier: Graduate School of Business, University of Chicago, Chicago, Illinois 60637.
- Reward Structure and Cooperation: The Decomposed Prisoner's Dilemma Game: Dean Pruitt*: Department of Psychology, State University of New York, Buffalo, New York 14214.
- Effects of Cognitive Desensitization on Avoidance Behavior: Stuart Valins* and Alice Allen: Department of Psychiatry, School of Medicine, University of North Carolina, Chapel Hill, North Carolina 27514.
- Relative Influence of Positive and Negative Information in Impression Formation and Persistence: Marjorie H. Richey,* Lucille McClelland, and Algimantas M. Shimkunas: Department of Psychology, Saint Louis University, 221 North Grand Boulevard, St. Louis, Missouri 63103.

* Asterisk indicates author for whom address is supplied.

PERSONAL RELIGIOUS ORIENTATION AND PREJUDICE

GORDON W. ALLPORT AND J. MICHAEL ROSS

Harvard University

3 generalizations seem well established concerning the relationship between subjective religion and ethnic prejudice: (a) On the average churchgoers are more prejudiced than nonchurchgoers; (b) the relationship is curvilinear; (c) people with an extrinsic religious orientation are significantly more prejudiced than people with an intrinsic religious orientation. With the aid of a scale to measure extrinsic and intrinsic orientation this research confirmed previous findings and added a 4th: people who are indiscriminately pro-religious are the most prejudiced of all. The interpretations offered are in terms of cognitive style.

Previous psychological and survey research has established three important facts regarding the relationship between prejudiced attitudes and the personal practice of religion.

1. On the average, church attenders are more prejudiced than nonattenders.

2. This overall finding, if taken only by itself, obscures a curvilinear relationship. While it is true that most attenders are more prejudiced than nonattenders, a significant minority of them are less prejudiced.

3. It is the casual, irregular fringe members who are high in prejudice; their religious motivation is of the extrinsic order. It is the constant, devout, internalized members who are low in prejudice; their religious motivation is of the *intrinsic* order.

The present paper will establish a fourth important finding—although it may properly be regarded as an amplification of the third. *The finding is that a certain cognitive style permeates the thinking of many people in such a way that they are indiscriminately proreligious and, at the same time, highly prejudiced.*

But first let us make clear the types of evidence upon which the first three propositions are based and examine their theoretical significance.

CHURCHGOERS ARE MORE PREJUDICED

Beginning the long parade of findings demonstrating that churchgoers are more intolerant of ethnic minorities than nonattenders is a study by Allport and Kramer (1946). These authors discovered that students who claimed no religious affiliation were less likely to be anti-Negro than those who de-

clared themselves to be protestant or Catholic. Furthermore, students reporting a strong religious influence at home were higher in ethnic prejudice than students reporting only slight or no religious influence. Rosenblith (1949) discovered the same trend among students in South Dakota. *The Authoritarian Personality* (Adorno, Frenkel-Brunswick, Levinson, & Sanford, 1950, p. 212) stated that scores on ethnocentrism (as well as on authoritarianism) are significantly higher among church attenders than among nonattenders. Gough's (1951) findings were similar. Kirkpatrick (1949) found religious people in general to be slightly less humanitarian than nonreligious people. For example, they had more punitive attitudes toward criminals, delinquents, prostitutes, homosexuals, and those in need of psychiatric treatment. Working with a student population Rokeach (1960) discovered nonbelievers to be consistently less dogmatic, less authoritarian, and less ethnocentric than believers. Public-opinion polls (as summarized by Stember, 1961) revealed confirmatory evidence across the board.

Going beyond ethnic prejudice, Stouffer (1955) demonstrated that among a representative sample of American church members those who had attended church within the past month were more intolerant of non-conformists (such as socialists, atheists, or communists) than those who had not attended. It seems that on the average religious people show more intolerance in general—not only toward ethnic but also toward ideological groups.

Is this persistent relationship in any way

spurious? Can it be due, for example, to the factor of educational level? Many studies show that people with high education tend to be appreciably less prejudiced than people with low education. Perhaps it is the former group that less often goes to church. The reasoning is false. Sociological evidence has shown conclusively that frequent church attendance is associated with high socioeconomic status and with college education (Demerath, 1965). Furthermore, Stouffer's study found that the intolerant tendency among churchgoers existed only when educational level was held constant. Struening (1963), using as subjects only faculty members of a large state university (all highly educated), discovered that nonattenders were on the average less prejudiced than attenders. These studies assure us that the association between churchgoing and prejudice is not merely a spurious product of low education.

Turning to the theoretical implications of these findings, shall we say that religion in and of itself makes for prejudice and intolerance? There are some arguments in favor of such a conclusion, especially when we recall that certain powerful *theological* positions—those emphasizing revelation, election (chosen people), and theocracy (Allport, 1959, 1966)—have throughout history turned one religion against another. And among *sociological* factors in religion we find many that make for bigotry. One thinks of the narrow composition of many religious groups in terms of ethnic and class membership, of their pressure toward conformity, and of the competition between them (see Demerath, 1965; Lenski, 1961). It does seem that religion as such makes for prejudice.

And yet it is here that we encounter the grand paradox. One may not overlook the teachings of equality and brotherhood, of compassion and humanheartedness, that mark all the great world religions. Nor may one overlook the precept and example of great figures whose labors in behalf of tolerance were and are religiously motivated—such as Christ himself, Tertullian, Pope Gelasius I, St. Ambrose, Cardinal Cusa, Sebastian Castellio, Schwenckfeld, Roger Williams, Mahatma Gandhi, Martin Luther King, and many others, including the recently martyred

clergy in our own South. These lives, along with the work of many religious bodies, councils, and service organizations would seem to indicate that religion as such *unmakes* prejudice. A paradox indeed.

THE CURVILINEAR RELATIONSHIP

If religion as such made *only* for prejudice, we would expect that churchgoers who expose themselves most constantly to its influence would, as a result, be more prejudiced than those who seldom attend. Such is not the case.

Many studies show that frequent attenders are less prejudiced than infrequent attenders and often less prejudiced even than nonattenders. Let us cite one illustrative study by Struening (1963). The curvilinear trend is immediately apparent in Table 1. In this particular study nonattenders had lower prejudice scores than any group, save only those devotees who managed to attend 11 or more times a month. Without employing such fine time intervals other studies have shown the same curvilinear trend. Thus, in *The Authoritarian Personality* (p. 212) we learned that in 12 out of 15 groups "regular" attenders (like nonattenders) were less prejudiced than "seldom" or "often" attenders. Employing a 26-item Desegregation Scale in three separate studies, Holtzman (1956) found the same trend as shown in Table 2. If more evidence for the curvilinear relationship is needed, it will be found in community studies made in New Jersey (Friedrichs, 1959), North Carolina (Tumin, 1958), New England (Pettigrew, 1959), and Ohio

TABLE 1
CHURCH ATTENDANCE AND PREJUDICE AMONG
FACULTY MEMBERS OF A MIDWESTERN
UNIVERSITY

Frequency of attendance (times per mo.)	N	Prejudice score
0	261	14.7
1	143	25.0
2	103	26.0
3	84	23.8
4	157	22.0
5-7	94	19.9
8-10	26	16.3
11 or more	21	11.7

Note.—From Struening (1957).

TABLE 2

CHURCH ATTENDANCE AND PREJUDICE AMONG STUDENTS
IN THE BORDER STATES

	1956 study % intolerant	Mean score on D scale	
		1958 study	1960 study
Nonattenders	37	41.3	38.1
Once a mo.	66	48.5	51.4
Twice a mo.	67	50.6	48.4
Once a wk. or oftener	49	44.5	44.3

Note.—Adapted from Holtzman (1956), Kelley, Ferson, and Holtzman (1958), Young, Benson, and Holtzman (1960).

and California (Pinkney, 1961). One could almost say there is a unanimity of findings on this matter. The trend holds regardless of religion, denomination, or target of prejudice (although the case seems less clear for anti-Semitism than for prejudice against other ethnic groups).

What are the theoretical implications? To find that prejudice is related to frequency of church attendance is scarcely explanatory, since it may reflect only formal behavior, not involvement or commitment to religious values. And yet it seems obvious that the regular attenders who go to church once a week or oftener (and several studies indicate that oftener than once a week is especially significant) are people who receive something of special ideological and experiential meaning. Irregular, casual fringe members, on the other hand, regard their religious contacts as less binding, less absorbing, less integral with their personal lives.

At this point, therefore, we must pass from external behavioral evidence into the realm of experience and motivation. Unless we do so we cannot hope to understand the curvilinear relationship that has been so clearly established.

EXTRINSIC VERSUS INTRINSIC MOTIVATION

Perhaps the briefest way to characterize the two poles of subjective religion is to say that the extrinsically motivated person *uses* his religion, whereas the intrinsically motivated *lives* his religion. As we shall see later, most people, if they profess religion at all, fall upon a continuum between these two poles. Seldom, if ever, does one encounter a

"pure" case. And yet to clarify the dimension it is helpful to characterize it in terms of the two ideal types.

Extrinsic Orientation

Persons with this orientation are disposed to use religion for their own ends. The term is borrowed from axiology, to designate an interest that is held because it serves other, more ultimate interests. Extrinsic values are always instrumental and utilitarian. Persons with this orientation may find religion useful in a variety of ways—to provide security and solace, sociability and distraction, status and self-justification. The embraced creed is lightly held or else selectively shaped to fit more primary needs. In theological terms the extrinsic type turns to God, but without turning away from self.

Intrinsic Orientation

Persons with this orientation find their master motive in religion. Other needs, strong as they may be, are regarded as of less ultimate significance, and they are, so far as possible, brought into harmony with the religious beliefs and prescriptions. Having embraced a creed the individual endeavors to internalize it and follow it fully. It is in this sense that he *lives* his religion.

A clergyman was making the same distinction when he said,

Some people come to church to thank God, to acknowledge His glory, and to ask His guidance. . . . Others come for what they can get. Their interest in the church is to run it or exploit it rather than to serve it.

Approximate parallels to these psychological types have been proposed by the sociologists Fichter (1954) and Lenski (1961). The former, in studying Catholic parishioners, classified them into four groups: the dormant, the marginal, the modal, and the nuclear. Omitting the dormant, Fichter estimated in terms of numbers that 20% are marginal, 70% modal, and less than 10% nuclear. It is, of course, the latter group that would most closely correspond to our conception of the "intrinsic." Lenski distinguished between church members whose involvement is "communal" (for the purpose of sociability and

status) and those who are "associational" (seeking the deeper values of their faith).

These authors see the significance of their classifications for the study of prejudice. Fichter has found less prejudice among devout (nuclear) Catholics than among others (see Allport, 1954, p. 421). Lenski (1961, p. 173) reported that among Detroit Catholics 59% of those with a predominantly "communal" involvement favored segregated schools, whereas among those with predominantly an "associational" involvement only 27% favored segregation. The same trend held for Detroit Protestants.

The first published study relating the extrinsic-intrinsic dimension directly to ethnic prejudice was that of Wilson (1960). Limiting himself to a 15-item scale measuring an extrinsic (utilitarian-institutional) orientation, Wilson found in 10 religious groups a median correlation of .65 between his scale and anti-Semitism. In general these correlations were higher than he obtained between anti-Semitism and the Religious-Conventionalism Scale (Levinson, 1954). From this finding Wilson concluded that orthodoxy or fundamentalism is a less important factor than extrinsicness of orientation.

Certain weaknesses may be pointed out in this pioneer study. Wilson did not attempt to measure intrinsicness of orientation, but assumed without warrant that it was equivalent to a low score on the extrinsic measures. Further, since the items were worded in a unidirectional way there may be an error of response set. Again, Wilson dealt only with Jews as a target of prejudice, and so the generality of his finding is not known.

Finally, the factor of educational level plays a part. Wilson used the California Anti-Semitism scale, and we know that high scores on this scale go with low education (Christie, 1954; Pettigrew, 1959; Titus & Hollander, 1957; Williams, 1964). Further, in our own study the extrinsic subscale is negatively correlated with degree of education ($r = -.32$). To an appreciable extent, therefore, Wilson's high correlations may be "ascribed" to educational level.

At this point, however, an important theoretical observation must be made. Low education may indeed predispose a person toward

an exclusionist, self-centered, extrinsic, religious orientation and may dispose him to a stereotyped, fearful image of Jews. This fact does not in the least affect the functional relationship between the religious and the prejudiced outlooks. It is a common error for investigators to "control for" demographic factors without considering the danger involved in doing so. In so doing they are often obscuring and not illuminating the functional (i.e., psychological) relationships that obtain (see Allport, 1950).

Following Wilson the task of direct measurement was taken up by Feagin (1964) who used a more developed scale—one designed to measure not only extrinsic orientation but also the intrinsic. His scales are essentially the same as those discussed in a later section of this paper. In his study of Southern Baptists Feagin reached four conclusions: (a) Contrary to expectation, extrinsic and intrinsic items did not fall on a unidimensional scale but represented two independent dimensions; (b) only the extrinsic orientation was related to intolerance toward Negroes; (c) orthodoxy as such was not related to the extrinsic or intrinsic orientation; (d) greater orthodoxy (fundamentalism of belief) did, however, relate positively to prejudice.

Taking all these studies together we are justified in assuming that the inner experience of religion (what it means to the individual) is an important causal factor in developing a tolerant or a prejudiced outlook on life.

Yet, additional evidence is always in place, and new insights can be gained by a closer inspection of the rather coarse relationships that have been established up to now.

THE PRESENT STUDY

We wished to employ an improved and broader measure of prejudice than had previously been used. And since direct measures of prejudice (naming the target groups) have become too sensitive for wide use, we wished to try some abbreviated indirect measures. Further, we wished to make use of an improved Extrinsic-Intrinsic scale, one that would give reliable measures of both extrinsic and intrinsic tendencies in a person's reli-

gibus life. For these reasons the following instruments were adopted.

Social Problems Questionnaire

This scale, devised by Harding and Schuman (unpublished¹; see also Schuman & Harding, 1963, 1964), is a subtly worded instrument containing 12 anti-Negro, 11 anti-Jewish, and 10 anti-other items (pertaining to Orientals, Mexicans, and Puerto Ricans). The wording is varied so as to avoid an agreement response set.

Indirect Prejudice Measures

Six items were taken from Gilbert and Levinson's (1956) Custodial Mental Illness Ideology Scale (CMI). Example: "We should be sympathetic with mental patients, but we cannot expect to understand their odd behavior. a) I definitely disagree. b) I tend to disagree. c) I tend to agree. d) I definitely agree."

Four items are related to a "jungle" philosophy of life, suggesting a generalized suspiciousness and distrust. Example: "The world is a hazardous place in which men are basically evil and dangerous. a) I definitely disagree. b) I tend to disagree. c) I tend to agree. d) I definitely agree."

In all cases the most prejudiced response receives a score of 5 and the least prejudiced response, 1. No response was scored 3.

From Table 3 we see that while the indirect measures have a positive correlation with each other and with direct measures the relationship is scarcely high enough to warrant the substitution of the indirect for the direct. The high correlations between prejudice for the three ethnic target groups

¹ J. Harding and H. Schuman, "Social Problems Questionnaire," Cornell University.

once again illustrate the well-established fact that ethnic prejudice tends to be a broadly generalized disposition in personality.

Religious Orientation Measure

The full scale, entitled "Religious Orientation," is available from ADI.² It separates the intrinsically worded items from the extrinsic, gives score values for each item, and reports on item reliabilities. In all cases a score of 1 indicates the most intrinsic response, a score of 5, the most extrinsic. While it is possible to use all 20 items as one continuous scale, it will soon become apparent that it is often wise to treat the two subscales separately. A sample item from the extrinsic subscale follows: "What religion offers me most is comfort when sorrows and misfortune strike. a) I definitely disagree, 1. b) I tend to disagree, 2. c) I tend to agree, 4. d) I definitely agree, 5." A sample item from the intrinsic subscale: "My religious beliefs are what really lie behind my whole approach to life. a) this is definitely not so, 5. b) probably not so, 4. c) probably so, 2. d) definitely so, 1.

SAMPLE

While our sample of six groups of churchgoers shows some diversity of denomination and region, it is in no sense representative. Graduate-student members of a seminar collected the 309 cases from the following church groups: Group A, 94 Roman Catholic (Massachusetts); Group B, 55 Lutheran (New York State); Group C, 44 Nazarene (South Carolina); Group D, 53 Presbyterian (Pennsylvania); Group E, 35 Methodist (Tennessee); Group F, 28 Baptist (Massachusetts).

We labeled the groups alphabetically since such small subsamples could not possibly lead to valid generalizations concerning denominations as a whole. All subjects knew that they were invited to participate as members of a religious group, and this fact may well have introduced a "proreligious" bias.

GROSS RESULTS

If we pool all our cases for the purpose of correlating religious orientation with preju-

TABLE 3
INTERCORRELATIONS BETWEEN FIVE MEASURES
OF PREJUDICE

	Anti-Jewish	Anti-Other	Jungle	CMI
Anti-Negro	.63	.70	.20	.25
Anti-Jewish		.67	.24	.31
Anti-Other			.33	.36
Jungle				.43

Note.—N = 309.

² The full Religious Orientation scale has been deposited with the American Documentation Institute. Order Document No. 9268 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

TABLE 4
CORRELATIONS BETWEEN EXTRINSIC SUBSCALE
AND PREJUDICE

Anti-Negro	.26
Anti-Jewish	.21
Anti-Other	.32
Jungle	.29
CMI	.44

Note.— $N = 309$.

dice, we discover that while the findings are in the expected direction they are much less impressive than those of previous studies, especially Wilson's.

Correlations with Extrinsic Subscale

Since Wilson employed an extrinsic scale similar to ours, we first present in Table 4 our findings using this subscale and the various measures of prejudice. Whereas Wilson found a correlation of .65 between his extrinsic and anti-Semitic measures, our correlation falls to .21. In part the reason no doubt lies in certain features of Wilson's method which we have criticized.

Correlations with Combined Extrinsic-Intrinsic Scale

From the outset it was our intention to broaden Wilson's unidirectional (extrinsic) measure to see whether our hypothesis might hold for the total scale (combined scores for the 11 extrinsic and 9 intrinsic items). As Table 5 shows, matters do not improve but seem to worsen. The logic of combining the two subscales is of course to augment the continuum in length and presumably enhance the reliability of the total measure. It soon became apparent, however, that subjects who endorse extrinsically worded items do not necessarily reject those worded intrinsically, or vice versa. It turns out that there is only

TABLE 5
CORRELATIONS BETWEEN TOTAL EXTRINSIC-
INTRINSIC SCALE AND PREJUDICE

Anti-Negro	.26
Anti-Jewish	.18
Anti-Other	.18
Jungle	.21
CMI	.17

Note.— $N = 309$.

a very low correlation in the expected direction between the two subscales ($r = .21$). Obviously at this point some reformulation is badly needed.

REFORMULATION OF THE APPROACH

Examination of the data reveals that some subjects are indeed "consistently intrinsic," having a strong tendency to endorse intrinsically worded items and to reject the extrinsically worded. Correspondingly others are "consistently extrinsic." Yet, unfortunately for our neat typology, many subjects are provokingly inconsistent. They persist in endorsing any or all items that to them seem favorable to religion in any sense. Their responses, therefore, are "indiscriminately pro-religious."

The problem is essentially the same as that encountered by the many investigators who have attempted to reverse the wording of items comprising the F scale, in order to escape an unwanted response-set bias. Uniformly the effort has proved to be frustrating, since so many subjects subscribe to both the positive and negative wording of the same question (see Bass, 1955; Chapman & Bock, 1958; Chapman & Campbell, 1959; Christie, 1954; Jackson & Messick, 1957).

An example from our own subscales would be: "My religious beliefs are what really lie behind my whole approach to life" (intrinsic). "Though I believe in my religion, I feel there are many more important things in my life" (extrinsic).

The approach used by Peabody (1961) offers us a model for analyzing our data in a meaningful way. Peabody administered both positive and negative F-scale items to subjects at two different testing sessions. By comparing each individual's responses to the same question stated positively at one time and in reverse at another he was able to separate out those who were consistently pro or anti toward the content of authoritarian items. But he found many who expressed double agreement (or disagreement) with both versions of the same question. Table 6 applies Peabody's paradigm to our data.

In assigning our 309 cases to these categories we employed the following criteria.

Intrinsic type includes individuals who

TABLE 6

FOUR PATTERNS OF RELIGIOUS ORIENTATION

	Agrees with intrinsic choice	Disagrees with intrinsic choice
Agrees with extrinsic choice	Indiscriminately proreligious	Consistently extrinsic in type
Disagrees with extrinsic choice	Consistently intrinsic in type	Indiscriminately antireligious or nonreligious ^a

^a Not found in present sample.

agree with intrinsically worded items on the intrinsic subscale, and who disagree with extrinsically stated items on the extrinsic subscale. By the scoring method employed these individuals fall below the median scores on both subscales.

Extrinsic type includes individuals who agree with extrinsically^a stated items on the extrinsic subscale, and who disagree with items on the intrinsic subscale. By our scoring method these individuals all fall above the median scores on both subscales.

Indiscriminately proreligious includes those who on the intrinsic subscale score at least 12 points less than on the extrinsic subscale. (This figure reflects the fact that a subject gives approximately 50% more intrinsic responses on the intrinsic subscale than we should expect from his extrinsic responses to the extrinsic subscale.)^a

Indiscriminately antireligious or nonreligious includes those who would show a strong tendency to disagree with items on both subscales. Since nonchurchgoers are excluded from our samples, such cases are not found. (Some pilot work with markedly liberal groups indicates that this type does exist, however, even among members of "religious" organizations.)

Table 7 gives the percentage of the three types.

RESULTS OF THE REFORMULATION

The five measures of prejudice were analyzed by a 6 (Groups) \times 3 (Religious Types) analysis of variance. Table 8 presents the overall effects for religious types for each of the five measures of prejudice. The multivariate analysis of variance indicates

TABLE 7

PERCENTAGE OF EACH RELIGIOUS TYPE IN EACH SUBSAMPLE

Religious group	N	Consistently intrinsic	Consistently extrinsic	Indiscriminately proreligious
A	(94)	36	34	30
B	(55)	35	36	29
C	(44)	36	39	25
D	(53)	32	30	38
E	(35)	31	29	40
F	(28)	39	39	22

that there is both a significant difference between the three types of religious orientation and between the six subsamples in the level of prejudice.³ Examination of the means shows two trends: (a) The extrinsic type is more prejudiced than the intrinsic type for both direct and indirect measures; (b) the indiscriminate type of religious orientation is more prejudiced than either of the two consistent types. Statistically all these trends are highly significant.

³ The multivariate *F* reported here is Wilk's lambda (Anderson, 1958). Statistical computations are summarized by Bock (1963) and programmed for the IBM 7090 by Hall and Cramer (1962). The univariate tests to be reported are adjusted for unequal *N*s to obtain orthogonal estimates according to mathematical procedures described in Hall and Cramer.

TABLE 8

PREJUDICE AND RELIGIOUS ORIENTATION

Target of prejudice	Mean prejudice score			<i>F</i> ratio
	Intrinsic type <i>N</i> = 108	Extrinsic type <i>N</i> = 106	Inconsistent type <i>N</i> = 95	
Anti-Negro	28.7	33.0	36.0	8.6**
Anti-Jewish	22.6	24.6	28.9	11.1**
Anti-Other	20.4	23.3	26.1	10.9**
Jungle	7.9	8.7	9.6	8.4**
CMI	10.2	11.8	13.4	20.4**

Multivariate analysis of variance

Source of variation	<i>F</i> ratio	<i>df</i>
Religious type (A)	5.96***	10,574
Sample groups (B)	3.19***	25,668
A \times B	1.11*	50,1312

* *p* > .25.** *p* > .001.*** *p* > .0005.

We note especially that the scores of the indiscriminate type are markedly higher on all measures than the scores of the intrinsic type. Corresponding *F* ratios for paired comparisons range from 8.4 for the jungle scale to 20.4 for the CMI scale. The differences between the indiscriminate and extrinsic types are smaller. For the anti-Jewish and CMI scales these differences are, however, beyond the .005 level; for the anti-other and jungle scales, at the .05 level. For the anti-Negro the difference falls below significance.

The relationship between the indiscriminately proreligious orientation and prejudice receives support (see Table 9) when we compare subjects who are *moderately* indiscriminate with those who are *extremely* indiscriminate. (In the first group the scores on the intrinsic subscale average 16 points lower than on the extrinsic subscale, whereas the extreme cases average 23 points less on the intrinsic than on the extrinsic subscale.)

The discovery that the degree of indiscriminateness tends to relate directly to the degree of prejudice is an important finding. It can only mean that some functional relationship obtains between religious muddle-headedness (for that is what indiscriminate scores imply) and antagonism toward ethnic groups. We shall return to this interpretation in the concluding section of this paper.

RESULTS FOR SUBSAMPLES

It would not be correct to assume that the variance is distributed equally over all the subsamples, for it turns out that the denominational groups differ appreciably in prejudice scores and in religious type, as Tables 10 and 11 indicate.

TABLE 9
DEGREES OF INDISCRIMINATENESS AND AVERAGE PREJUDICE SCORES

Target of prejudice	Moderately indiscriminate <i>N</i> = 56	Extremely indiscriminate <i>N</i> = 39	<i>F</i> ratio
Anti-Negro	35.4	37.9	.97
Anti-Jewish	28.0	30.1	.90
Anti-Other	24.9	28.2	3.25*
Jungle	9.5	10.2	1.11
CMI	10.2	14.6	3.99*

**p* > .05.

TABLE 10
ANTI-NEGRO PREJUDICE: MEAN SCORES ON SOCIAL PROBLEMS SCALE

Religious group	Intrinsic type	Extrinsic type	Indiscriminate type	Group <i>M</i>
A	27.4 (34)	34.8 (32)	32.2 (28)	31.4 (94)
B	27.2 (19)	32.3 (20)	31.9 (16)	30.4 (55)
C	22.4 (16)	36.2 (17)	35.0 (11)	30.9 (44)
D	35.5 (17)	28.7 (16)	42.5 (20)	36.1 (53)
E	40.5 (11)	35.5 (10)	43.0 (14)	40.1 (35)
F	22.6 (11)	27.9 (11)	28.7 (6)	26.0 (28)
Type <i>M</i>	28.7 (108)	33.0 (106)	36.0 (95)	32.5 (309)

Analysis of variance

Source of variation	<i>df</i>	<i>MS</i>	<i>F</i> ratio
Religious type (A)	2	1077.8	8.6**
Religious group (B)	5	952.2	7.6**
A × B	10	251.1	2.0*
Error (w)	291	125.6	

**p* > .10.
***p* > .001.

It is true that when we combine subsamples all the trends are in the expected direction, but troublesome exceptions occur for single groups as indicated by the nearly significant interaction effects. The most troublesome contradictions appear in relation to the anti-Negro measures based on the Harding-Schuman scale. Table 10 discloses certain sore points, even though the average trend over all the subsamples is in the predicted direction.

For Groups A, B, and C we note that the indiscriminate type is slightly less prejudiced than the extrinsic type, and for Groups D and E the extrinsic type seems actually less prejudiced than the intrinsic. (Groups D and E are consistently more troublesome than other subsamples, perhaps because of some salient racial issue in the local community. It will be noted that both these groups are considerably more anti-Negro than the other subsamples.)

By way of contrast we present in Table 11 the results for the short (five-item) CMI scale. With the exception of the indiscriminate type in Group F, the progression of scores is precisely as expected. Each subsample shows that the intrinsic type is less prejudiced toward the mentally ill than the extrinsic type, and the extrinsic type is less

TABLE 11
INDIRECT (CMI) MEASURE OF PREJUDICE

Religious group	Intrinsic type	Extrinsic type	Indiscriminate type	Group M
A	11.2 (34)	12.4 (32)	13.6 (28)	12.3 (94)
B	10.1 (19)	10.8 (20)	13.4 (16)	11.3 (55)
C	9.5 (16)	12.2 (17)	12.6 (11)	11.3 (44)
D	10.6 (17)	11.4 (16)	14.8 (20)	12.4 (53)
E	8.6 (11)	12.9 (10)	13.6 (14)	11.8 (35)
F	9.2 (11)	10.7 (11)	9.2 (6)	9.8 (28)
Type M	10.2 (108)	11.8 (106)	13.4 (95)	11.9 (309)

Analysis of variance

Source of variation	df	MS	F ratio
Religious type (A)	2	255.0	20.4**
Religious group (B)	5	36.5	2.9*
A × B	10	15.3	1.2
Error (w)	291	12.5	

* $p > .05$.** $p > .001$.

prejudiced than the indiscriminately prejudiced.⁴

Returning in a different way to the original question of whether consistent extrinsic and intrinsic orientations make for prejudice and for tolerance, respectively, we shall now examine this matter in each subsample separately. Inspection of the mean scores and

⁴ If we apply a more severe test, asking whether all differences between groups are significant, we find the following results. In four of the six groups (in both Tables 10 and 11) the extrinsic type is significantly more prejudiced than the intrinsic. Likewise in four out of six groups (Table 10) and five out of six (Table 11), the indiscriminate type is significantly more prejudiced than the intrinsic. However, in only two of the six groups (in both Tables 10 and 11) is the indiscriminate type significantly more prejudiced than the extrinsic.

variance for the total scale indicates that we are dealing with a relatively narrow range of variation. To minimize the effect of a narrow range of scores and skewed distributions, we used Kendal's (1955) tau as a measure of degree of relationship between prejudice and consistent religious orientation. The results are given in Table 12. While the correlations are not high (14 are significant in the expected direction), only one (in the troublesome Group E) is significant in the reverse direction.

EDUCATIONAL DIFFERENCES

Computing the actual years of schooling for all groups we find that the indiscriminate type has significantly less formal education than the intrinsic cases ($p > .005$, $F = 18.29$), and somewhat less than the extrinsic type ($p > .10$, $F = 2.89$). Comparing extrinsic with intrinsic types we find that the former has finished fewer years of schooling ($p > .10$, $F = 3.45$). (Oddly enough the groups with highest average education are D and E, which also displayed the highest anti-Negro and anti-Semitic prejudice—perhaps because of particular local conditions.)

In our survey of earlier studies we saw that educational level is often a factor in the various relationships discovered between religion and prejudice. We have also argued that demographic factors of this sort should not be allowed to obscure the functional (psychological) analysis that the data call for. Granted that low education makes for indiscriminate thinking, the mental confusion that results from low education may have its own peculiar effects on religious and ethnic attitudes.

TABLE 12
CORRELATIONS BETWEEN COMBINED EXTRINSIC-INTRINSIC RELIGIOUS SCORES (FOR CONSISTENT SUBJECTS) AND PREJUDICE (KENDAL'S TAU)

Religious group	Anti-Negro	Anti-Jewish	Anti-Other	Jungle	CMI
A	.31***	.26***	.24***	.14*	.19***
B	.19*	.13	.15	-.05	.03
C	.32***	.17*	.35***	.14*	.28***
D	-.12	.05	-.09	.03	.11
E	-.24*	-.11	-.13	.26*	.46***
F	.39***	.13	.25*	-.01	.24*

* $p > .10$.*** $p > .05$.*** $p > .01$.

SUMMARY AND INTERPRETATIONS

At the outset we stated three propositions that seem to be firmly established: (a) Churchgoers on the broad average harbor more ethnic prejudice than nonchurchgoers; (b) in spite of this broad tendency a curvilinear relationship in fact exists; (c) the intrinsically motivated churchgoers are significantly less prejudiced than the extrinsically motivated. Our present research supplies additional strong support for the second and third of these propositions.

To these propositions we add a fourth: *churchgoers who are indiscriminately pro-religious are more prejudiced than the consistently extrinsic, and very much more prejudiced than the consistently intrinsic types.*

The psychological tie between the intrinsic orientation and tolerance, and between the extrinsic orientation and prejudice, has been discussed in a series of papers by Allport (1959, 1963, 1966). In brief the argument holds that a person with an extrinsic religious orientation is using his religious views to provide security, comfort, status, or social support for himself—religion is not a value in its own right, it serves other needs, and it is a purely utilitarian formation. Now prejudice too is a “useful” formation: it too provides security, comfort, status, and social support. A life that is dependent on the supports of extrinsic religion is likely to be dependent on the supports of prejudice, hence our positive correlations between the extrinsic orientation and intolerance. Conversely, the intrinsic religious orientation is not an instrumental device. It is not a mere mode of conformity, nor a crutch, nor a tranquilizer, nor a bid for status. All needs are subordinated to an overarching religious commitment. In internalizing the total creed of his religion the individual necessarily internalizes its values of humility, compassion, and love of neighbor. In such a life (where religion is an intrinsic and dominant value) there is no place for rejection, contempt, or condescension toward one's fellow man. Such is our explanation for the relationship between extrinsic religion and prejudice, and between intrinsic religion and tolerance.

Our present task is to discover, if we can,

some similar functional tie between prejudice (as measured both directly and indirectly) and the indiscriminately proreligious orientation. The common factor seems to be a certain cognitive style. Technically it might be called “undifferentiated thinking,” or excessive “category width,” as defined by Pettigrew (1958). Rokeach (1960) notes the inability of the “dogmatic” mind to perceive differences; thus, whereas some people distinguish in their thinking and feeling between Communists and Nazis, the undifferentiated dogmatist has a global reaction (cognitive and emotional) toward “Communazis.”

We have no right, of course, to expect all our subjects to make discriminations exactly corresponding to our own logic. Nor should we expect them to read and respond to every item on the Extrinsic-Intrinsic scale according to its full meaning as intended by the investigators. Perhaps we should be gratified that two-thirds of our cases can be safely classified as “consistent” (i.e., having about the same strength of disposition toward an extrinsic or intrinsic orientation across most of the items). These consistent cases, as we have seen, support the hypothesis with which we started. It is the remaining (indiscriminate) one-third of the cases which obscure the trend (or diminish its statistical significance).

In responding to the religious items these individuals seem to take a superficial or “hit and run” approach. Their mental set seems to be “all religion is good.” “My religious beliefs are what really lie behind my whole life”—Yes! “Although I believe in my religion, I feel there are many more important things in my life”—Yes! “Religion is especially important to me because it answers many questions about the meaning of life”—Yes! “The church is most important as a place to formulate good social relationships”—Yes!

There seems to be one wide category—“religion is OK.” From the way in which the scale is constructed this undifferentiated endorsement can be the product of an agreement response set. Our inconsistently pro-religious may be “yeasayers” (Couch & Keniston, 1960). But if so, we are still dealing with an undifferentiated cognitive

disposition. We recall likewise that the inconsistent cases have a lower level of formal education than the consistent cases. This factor also is relevant to the formation and holding of worldwide categories.

But why should such a disposition, whatever its source, be so strongly related to prejudice, in such a way that the more undifferentiated, the more prejudiced as Table 9 shows?

The answer is that prejudice itself is a matter of stereotyped overgeneralization, a failure to distinguish members of a minority group as individuals (Allport, 1954, Chaps. 2, 10). It goes without saying that if categories are overwide the accompanying feeling tone will be undifferentiated. Thus, religion as a whole is good; a minority group as a whole is bad.

It seems probable that people with undifferentiated styles of thinking (and feeling) are not entirely secure in a world that for the most part demands fine and accurate distinctions. The resulting diffuse anxiety may well dispose them to grapple onto religion and to distrust strange ethnic groups. The positive correlation between the jungle items and other prejudice scales (Table 3) is evidence for this interpretation.

Our line of reasoning, readers will recognize, is compatible with various previous contributions to the theory of prejudice. One thinks here of Rokeach's concept of dogmatism; of Schuman and Harding's (1964) discovery of a "confused" type in their study of the relation between rational consistency and prejudice; of the same authors' work on sympathetic identification (1963); of studies on the dynamics of scapegoating, the role in insecurity; of authoritarian submission, of intolerance for ambiguity, and of related concepts.

All in all we conclude that prejudice, like tolerance, is often embedded deeply in personality structure and is reflected in a consistent cognitive style. Both states of mind are emeshed with the individual's religious orientation. One definable style marks the individual who is bigoted in ethnic matters and extrinsic in his religious orientation. Equally apparent is the style of those who are bigoted and at the same time indiscrimi-

nately proreligious. A relatively small number of people show an equally consistent cognitive style in their simultaneous commitment to religion as a dominant, intrinsic value and to ethnic tolerance.

One final word: our research is so strongly that social scientists who take the variable "religion" or "religiosity" in the future will do well to keep in mind the crucial distinction between religious attitudes that are *intrinsic*, *extrinsic*, and *indiscriminately pro*. To know that a person is in some sense "religious" is not as important as to know the role religion plays in the economy of his life. (The categories of *nonreligious* and *indiscriminately antireligious* will also for some purposes be of central significance, although the present research, confined as it is to churchgoers, does not employ them.)

REFERENCES

- ADORNO, T. W., FRENKEL-BRUNSWIK, E., LIVINGSTON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
- ALLPORT, G. W. Review of S. A. Stouffer, F. A. Suchman, L. C. De Vinney, S. A. Star, & R. W. Williams, Jr., *The American soldier*. Vol. 1. *Adjustment during Army life*. *Journal of Abnormal and Social Psychology*, 1950, 45, 168-173.
- ALLPORT, G. W. *The nature of prejudice*. Reading, Mass.: Addison-Wesley, 1954.
- ALLPORT, G. W. Religion and prejudice. *The Crisis Review*, 1959, 2, 1-10.
- ALLPORT, G. W. Behavioral science, religion, and mental health. *Journal of Religion and Health*, 1963, 2, 187-197.
- ALLPORT, G. W. Religious context of prejudice. *Journal for the Scientific Study of Religion*, 1966, 5, 447-457.
- ALLPORT, G. W., & KRAMER, B. M. Some roots of prejudice. *Journal of Psychology*, 1946, 22, 1-10.
- ANDERSON, T. W. *An introduction to multivariate statistical analysis*. New York: Wiley, 1958.
- BASS, B. M. Authoritarianism or acquiescence. *Journal of Abnormal and Social Psychology*, 1955, 56, 616-623.
- BOCK, R. D. Programming univariate and multivariate analysis of variance. *Technometrics*, 1960, 5, 95-117.
- CHAPMAN, L. J., & BOCK, R. D. Components of variance due to acquiescence and content in the F-scale measure of authoritarianism. *Psychological Bulletin*, 1958, 55, 328-333.
- CHAPMAN, L. J., & CAMPBELL, D. T. The effect of acquiescence response-set upon relationships among the F-scale, ethnocentrism, and intelligence. *Sociometry*, 1959, 22, 153-161.
- CHRISTIE, R. C. Authoritarianism re-examined. In R. C. Christie & M. Jahoda (Eds.), *Studies in the*

- scope and method of the authoritarian personality. New York: Free Press of Glencoe, 1954. Pp. 113-196.
- LOREN, A., & KENNEDY, K. Yemassee and maynards: Agreeing response set as a personality variable. *Journal of Abnormal and Social Psychology*, 1961, 60, 151-174.
- MACMURRAY, N. J., III. *Social class in American Protestantism*. Chicago: Rand McNally, 1965.
- PEACOCK, J. R. Prejudice and religious types: A focused study of southern fundamentalists. *Journal for the Scientific Study of Religion*, 1964, 4, 6-13.
- PICHTER, J. H. *Social relations in the urban parish*. Chicago: University of Chicago Press, 1954.
- PERDUE, R. W. Christians and residential exclusion: An empirical study of a northern dilemma. *Journal of Social Issues*, 1959, 15, 14-23.
- QUIGLEY, D. C., & LEVINSON, D. J. Ideology, personality, and institutional policy in the mental hospital. *Journal of Abnormal and Social Psychology*, 1956, 53, 263-271.
- GOVERN, H. G. Studies in social intolerance: IV. *Journal of Social Psychology*, 1951, 33, 263-269.
- HALL, C. E., & CRAMER, E. *General purpose program to compute multivariate analysis of variance on an IBM 7090*. Washington, D. C.: George Washington University Biometric Laboratory, 1962.
- HOLTZMAN, W. H. Attitudes of college men toward non-segregation in Texas schools. *Public Opinion Quarterly*, 1956, 20, 559-569.
- JACKSON, D. H., & MASSICK, S. J. A note on ethnocentrism and acquiescence response sets. *Journal of Abnormal and Social Psychology*, 1957, 54, 132-134.
- KELLY, J. G., FERNON, J. E., & HOLTZMAN, W. H. The measurement of attitudes toward the Negro in the South. *Journal of Social Psychology*, 1958, 48, 305-317.
- KENDALL, M. G. *Rank correlation methods*. (2nd ed.) London: Griffin, 1955.
- KIRKPATRICK, C. Religion and humanitarianism: A study of institutional implications. *Psychological Monographs*, 1949, 63(9, Whole No. 304).
- LENSKI, G. *The religious factor*. Garden City, N. Y.: Doubleday, 1961.
- LEVINSKY, D. J. The inter-group workshop: Its psychological aims and effects. *Journal of Psychology*, 1954, 38, 103-126.
- FRANCOY, D. Attitude content and agreement set in scales of authoritarianism, dogmatism, anti-semitism, and economic conservatism. *Journal of Abnormal and Social Psychology*, 1962, 62, 1-11.
- FRANCOY, W. T. F. The questionnaire and attitude: A review of the literature. *Journal of Personality*, 1953, 21, 1-10.
- FRANCOY, W. T. F. Cultural differences in self-concept. *Journal of Abnormal and Social Psychology*, 1955, 49, 16-22.
- FRANCOY, A. The structure of prejudice: Making sense of the social distance scale in social attitudes and prejudiced action. *Journal of Social Issues*, 1965.
- FRANCOY, M. Prejudice and social mental functioning: An analysis of social systems and personality. New York: Free Press, 1966.
- FRANCOY, W. T. F. A synthesis of three studies of prejudice. *Journal of Abnormal and Social Psychology*, 1961, 64, 47-52.
- SCHUMAN, H., & HARTMAN, J. Segregation, prejudice, and the suburbs. *Public Opinion Quarterly*, 1961, 25, 200-241.
- SCHUMAN, H., & HARTMAN, J. Prejudice and the norm of rationality. *Sociometry*, 1964, 27, 353-371.
- STOMMER, H. C. *Education and attitude change*. New York: Institute of Human Relations Press, 1962.
- STOUTER, S. A. *Communism, the atom, and conformity*. Garden City, N. Y.: Doubleday, 1957.
- STREIBING, E. L. Attitudes toward the Negro in a Midland region. In H. H. Postman (Ed.), *Anti-Semitism: A study in prejudice and Evaluation*. Northwestern University Press, 1961, Ch. 9.
- THUR, H. E., & HOLLANDER, E. P. The Goldberg-E scale in psychological research, 1955-1960. *Psychological Bulletin*, 1967, 54, 47-64.
- TURNER, M. *Disorientation: Resistance and resistance*. Princeton: Princeton University Press, 1958.
- WILLIAMS, R. M. *Seaside next door: Ethnic relations in American communities*. Englewood Cliffs, N. J.: Prentice-Hall, 1964.
- WILSON, W. C. Extrinsic religious values and prejudice. *Journal of Abnormal and Social Psychology*, 1960, 60, 286-288.
- YOUNG, R. K., BENSON, W. M., & HOLTZMAN, W. H. Changes in attitudes toward the Negro in a southern university. *Journal of Abnormal and Social Psychology*, 1960, 60, 131-133.

(Received July 6, 1966)

AFFECTIVE STIMULUS VALUE AND COGNITIVE COMPLEXITY¹

MARC IRWIN

University of California, Berkeley

TONY TRIPODI

University of Michigan

AND JAMES BIERI

University of Texas

The issue as to whether one differentiates more among persons who evoke negative affect than among persons who evoke positive affect was investigated in 2 studies. Using cognitive complexity as a measure of differentiation, results from both studies indicate significantly more differentiation among negative stimulus persons than among positive persons. These differences were particularly marked for female Ss. Further, in the 1st study, neutral persons were differentiated significantly less than negative persons but significantly more than positive persons. The findings are discussed in relation to a vigilance hypothesis of the effect of negative affect upon judgmental processes.

Do we differentiate more among persons whom we regard with negative affect than among those toward whom we feel quite positive? The answer to such an apparently simple question is not at all obvious and bears upon rather fundamental issues in personality and perception that warrant careful empirical analysis. Results from studies on need and perception are neither entirely consistent nor entirely relevant in this regard. Evidence can be garnered for similar perceptual effects in relation to both positive and negative stimuli (Allport, 1955; Postman, 1953). One problem has been that *subthreshold* effects have received undue attention, considering the fact that most of our judgments of others are in relation to *supra-threshold* stimuli. Further, perceptual discrimination effects have been studied often only in relation to comparison of neutral and positive stimuli, as in the work of Dukes and Bevan (1952).

Two competing conceptual analyses may guide our thinking in regard to this problem. From an adaptive point of view, one could argue that the individual should differentiate more finely among negative, potentially threatening figures in his social environment so as to be able to isolate and identify these

potentially dangerous individuals. Such a *vigilance* hypothesis has received some support in the study by Miller and Bieri (1965). These writers found that scores indicative of greater cognitive complexity were found when subjects judged five persons toward whom they were relatively more distant and toward whom they might be expected to relate negatively (e.g., "person you dislike," "person with whom you feel most uncomfortable," and "boss") than among those toward whom more positive effect might be expected (e.g., "mother," "friend of opposite sex," "friend of same sex"). However, these results may be questioned since the affective evaluation could be positive or negative for such persons as "mother" and "boss."

The alternative hypothesis would suggest that one could differentiate more among those with whom predominantly positive feelings have been engendered. It can be argued, for example, that positive persons are those one knows best from direct experience, those with whom one desires to associate, and with whom one has greater contact and understanding. Negative, anxiety-laden figures, by contrast, are those one tends to avoid and, through this lack of frequent, direct contact, opportunities for greater discrimination are lost or limited. Crockett (1965) has discussed the results of a study by Supnick which are consistent with this latter *avoidance* hypothesis. However, the method for measuring

¹ This research was supported by Grant 486, Institute of Social Sciences, University of California, Berkeley, and Grant GS-842 from the National Science Foundation.

cognitive complexity in that study was substantially different from that used in the Miller and Bieri study.

To investigate these apparently competing hypotheses more systematically, two independent studies were completed. Both studies use the Role Construct Repertory Test method (Kelly, 1955) for assessing cognitive complexity originally advanced by Bieri (1955) and subsequently modified by Tripodi and Bieri (1963). In addition, a further modification in this procedure was used in the second study to be reported, allowing for a comparison of findings with two differing methods for measuring cognitive complexity. As is evident in research to date (Bieri, 1965; Bonarius, 1965), such analysis of the equivalence of different measures of the same construct is a major need in current cognitive structure research.

METHOD

Study 1

Undergraduates volunteered from three fraternities and three sororities at the University of California, Berkeley, yielding a total of 64 male and 51 female subjects. To assess the degree of differentiation among others, the modification of the Role Construct Repertory Test procedure devised by Tripodi and Bieri (1963) was used. Each subject was first given a card on which were printed 12 numbered spaces and asked to write in order the initials of his four housemates whom he liked most in spaces 1-4. In spaces 5-8, he was asked to write the initials of the four housemates about whom he felt neutral, that is, neither liked nor disliked. Next, working backwards from Space 12 to Space 9, the subject was asked to list the initials in order of the four housemates whom he disliked most. Subjects were assured that their sociometric choices would be kept confidential.

Each of these 12 persons' initials was entered on a grid in counterbalanced order of blocks of four liked, neutral, and disliked persons. The order within blocks was randomized. Next, the subject rated all 12 persons on each of 10 provided construct dimensions using a 6-point scale. The 10 construct dimensions used were outgoing versus shy, adjusted versus maladjusted, decisive versus indecisive, calm versus excitable, interested in others versus self-absorbed, cheerful versus ill-humored, responsible versus irresponsible, considerate versus inconsiderate, independent versus dependent, and interesting versus dull. Further details of this grid procedure can be found in Bieri, Atkins, Briar, Leaman, Miller, and Tripodi (1966). Cognitive complexity scores, indicative of degree of differentiation, were generated by comparing judgments in each row with those on

the same person in all subsequent rows, such that a high score is indicative of less cognitive complexity.

Study 2

Forty undergraduate male students and 40 undergraduate female students at the University of Texas comprised the sample in the second study. Several changes in procedure from the first study were made in regard to the judgment task. Each subject judged four persons whom he liked or felt positively toward and four persons whom he disliked or felt negatively toward, as in Study 1 (no neutral persons were judged in Study 2). However, half of each of the persons judged were of the same sex as the subject and half were of the opposite sex, in contrast to Study 1 in which all persons judged were of the same sex as the subject. Further, each person was judged using a single stimulus presentation rather than the grid format of Study 1. That is, in Study 1 all persons were judged on the same construct dimension before proceeding to the next construct dimension. This grid format could facilitate comparisons by the subject of previous judgments with subsequent judgments. In Study 2, each judgment was a random pairing of one of the eight persons with one of the eight construct dimensions. In this manner, direct comparison among judgments of a person on different dimensions is minimized, making the task less one of comparative judgment than in Study 1.

The four positive persons judged included the same-sex and opposite-sex "closest friend" (or spouse) and the same-sex and opposite-sex friend "you admire." The four negative-affect persons judged were the same- and opposite-sex person "you find hard to like" and "with whom you feel most uncomfortable." The eight construct dimensions used, with the exception of independent versus dependent and interesting versus dull, were the same constructs as in Study 1. Because only eight persons were judged on eight construct dimensions in Study 2, raw scores of cognitive complexity will be lower than in Study 1.

RESULTS

Study 1

Table 1 presents the means and standard deviations for cognitive complexity scores in Study 1 for judgments of persons with posi-

TABLE 1
COGNITIVE COMPLEXITY SCORES FOR THREE
DEGREES OF AFFECT (STUDY 1)

	CC+		CCn		CC-	
	M	SD	M	SD	M	SD
Males (<i>N</i> = 64)	65.4	19.4	55.6	21.3	46.7	14.7
Females (<i>N</i> = 51)	72.5	22.7	53.8	19.3	41.9	10.3

TABLE 2

COGNITIVE COMPLEXITY SCORES FOR POSITIVE AND NEGATIVE PERSONS (STUDY 2)

	CC+		CC-	
	M	SD	M	SD
Males (N = 40)	43.3	16.0	34.0	9.8
Females (N = 40)	47.4	16.3	27.6	7.5

tive (CC+), negative (CC-), and neutral affect (CCn). The reader will recall that low scores reflect greater cognitive complexity. Using Scheffé's test for multiple comparisons (Edwards, 1960), it was found that for both male and female subjects, positive figures were differentiated significantly less well than negative figures ($p < .001$). Further, for both males and females, positive figures were less well differentiated than neutral figures ($p < .001$), and neutral figures were less differentiated than negative figures ($p < .001$, males; $p < .01$, females). From the means presented in Table 1, it can be observed that males tended to differentiate more in relation to positive figures than did females, although this difference is not significant ($t = 1.78$, $p < .10$). Conversely, females compared to males tended to differentiate more among neutral and negative figures, the latter sex difference being statistically significant ($t = 2.07$, $p < .05$).

Study 2

The same basic analyses from the second study are presented in Table 2, and a striking consistency with the results of the first study can be noted. As in Study 1, both male and female subjects differentiated significantly less when judging positive-affect persons than when judging negative-affect persons ($t = 3.69$, $p < .001$; $t = 8.27$, $p < .001$, respectively). Further, as in the first study, males tended to differentiate more than females among positive persons ($t = 1.13$, *ns*), while females differentiated significantly more than males among the negative persons ($t = 3.22$, $p < .01$).

DISCUSSION

The results of both studies unequivocally support the proposition that individuals dif-

ferentiate more among persons who evoke more negative affect than among persons with whom strong positive-regard tendencies are associated. While no direct evidence can be obtained from these studies as to the basis for such tendencies, if one assumes that negative affect is related to danger or threat, the results are consistent with the vigilance hypothesis advanced by Miller and Bieri (1965). The finding in Study 1 that neutral-affect persons were differentiated at an intermediate degree between positive and negative persons may support the supposition that as the threat value of another increases, the individual seeks to articulate more finely the other's behavior. It is possible that such heightened tendencies to differentiate persons with negative affect reflect an effort to articulate others who are threatening. Kelly (1955) has suggested there is a need to gain greater predictability and understanding of persons who evoke anxiety responses. Alternative explanations need to be explored, however. For example, it could be that negative affect is associated with greater stimulus clarity or saliency than is positive affect, rendering greater differentiation among negative persons more possible.

These results call into question, however, the hypothesis that we differentiate better among those persons with whom we feel most positive and with whom perhaps we have the most direct contact (Crockett, 1965). The results reported by Crockett which apparently run counter to those of the present studies can perhaps be understood by considering the nature of the task to assess cognitive complexity used by Supnick (Crockett, 1965). Essentially, the subject was asked to describe in writing each of a number of persons, and the number of interpersonal constructs in these free descriptions was taken as the index of cognitive complexity. It would seem reasonable, as Crockett reports, that subjects would perhaps write more and describe more fully persons about whom they felt more positive in comparison to negative persons. It should be noted, however, that such a task makes no assessment of how functionally different the constructs may be inasmuch as the subject is not required to actually use these dimensions to discriminate between in-

valid. Indeed, an assumption of the method used to assess cognitive complexity in the present studies and which has been used in various modifications in a large number of previous studies (Bieri et al., 1966) is that only a task requiring the subject to make discriminations about others using his own or provided construct dimensions can provide an index of the number of independent dimensions in his cognitive structure. The ability to generate semantic categories in itself is no guarantee of greater cognitive complexity.²

Such method considerations do underscore the necessity for caution in generalizing about the nature of a construct such as cognitive complexity when widely disparate assessment techniques are used. In the two studies reported here, while equivalent results were obtained from differing methods of stimulus presentation, the responses demanded of the subject and the analytic methods used to derive cognitive complexity scores were equivalent. At the same time, the present results also suggest that the nature of the persons being judged does influence cognitive complexity. In general, persons who are more highly differentiating in judging positive persons are also more highly differentiating in judging negative persons, but there are discrepancies among the samples in the two studies. In Study 1, among males the correlations between cognitive complexity scores were $r = .53$, ($p < .01$) for positive and negative persons, $r = .45$ ($p < .01$) for positive and neutral persons, and $r = .56$ ($p < .01$) for neutral and negative persons. In Study 2, the correlation between cognitive complexity scores for positive and negative persons was, $r = .34$ ($p < .05$). For females, the correlations were less consistent. Thus, in Study 2 the cognitive complexity scores for positive and negative persons correlated $r = .40$ ($p < .01$), while in Study 1 this correlation was $r = .04$. In Study 1, the correlation be-

tween complexity scores for positive and neutral persons was $r = -.09$, and the correlation between neutral and negative persons was $r = .37$ ($p < .01$).

More consistent, however, is the significant tendency in both studies for females relative to males to differentiate more among persons whom they regard negatively than among those they regard positively. It could be that these results reflect the generally accepted belief that women have greater needs to depersonalize others and the greater ability to differentiate among potentially threatening or unrelatable others in a negative value in this regard. Finally, we may note that while judgment of others who have strong positive and negative affective value for the individual would seem to be the kind of situation in which socially desirable response tendencies might make a difference, no strong evidence for the influence of this variable was found in the two studies reported here. For example, in Studies 1 and 2 it was observed that for both males and females, correlations between the Marlowe-Crowne Social Desirability scale and the various cognitive complexity scores ranged from $r = .04$ to $r = .18$, all nonsignificant correlations. It should be noted that all construct dimensions used in the present study had both a "good" pole and a "bad" pole (e.g., adjusted-maladjusted). Judgment dimensions with poles that were equally "good" or "bad" might produce differentiations among others that would be more influenced by a social desirability difference among subjects.

It is reasonable that the results of the present studies need fuller amplification in relation to several parameters. In particular it should be pointed out that subjects were judging others in a relatively nonaroused emotional state. The experimental instigation of greater affective arousal while judging negative stimuli, for example, might elicit less rather than more ability to differentiate among such stimuli. While such an outcome does seem reasonable, the paucity of consistent evidence on this issue remains a major void. Aside from general formulations stemming from such roots as the Yerkes-Dodson law (Broadhurst, 1957; Easterbrook, 1959), the systematic effects of levels of af-

² Evidence for this assertion was obtained on a sample of 20 subjects. The number of constructs from the free-response method of Crockett and the cognitive complexity grid procedure correlated $r = .12$. That is, to a nonsignificant degree, less cognitively complex subjects generated more verbal categories.

reactive arousal upon discrimination behavior, for example are in need of more careful mapping. Given the fact that our perceptions of others are intimately associated with strong affects both positive and negative, further effort of this sort is clearly called for.

REFERENCES

- ALLPORT, F. H. *Theories of perception and the concept of structure*. New York: Wiley, 1955.
- BIERI, J. Cognitive complexity-simplicity and predictive behavior. *Journal of Abnormal and Social Psychology*, 1955, **51**, 263-268.
- BIERI, J. Cognitive complexity: Assessment issues in the study of cognitive structure. In J. I. Bardon (Chm.), *The cognitive approach to personality assessment*. Symposium presented at American Psychological Association, Chicago, September 4, 1965.
- BIERI, J., ATKINS, A. L., BRIAR, S., LEAMAN, R. L., MILLER, H., & TRIPODI, T. *Clinical and social judgment: The discrimination of behavioral information*. New York: Wiley, 1966.
- BONARIUS, J. C. J. Research in the personal construct theory of George A. Kelly: Role Construct Repertory Test and basic theory. In B. A. Maher (Ed.), *Progress in experimental personality research*. Vol. 2. New York: Academic Press, 1965. Pp. 1-46.
- BROADBENT, P. L. Emotionality and the Yerkes-Dodson law. *Journal of Experimental Psychology*, 1957, **54**, 545-552.
- CROCKETT, W. H. Cognitive complexity and impression formation. In B. A. Maher (Ed.), *Progress in experimental personality research*. Vol. 2. New York: Academic Press, 1965. Pp. 47-90.
- DUKES, W. F., & BEVAN, W. Accentuation and response variability in the perception of personally relevant objects. *Journal of Personality*, 1952, **20**, 457-465.
- EASTERBROOK, J. A. The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, 1959, **66**, 183-201.
- EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1960.
- KELLY, G. A. *The psychology of personal constructs*. Vol. 1. New York: Norton, 1955.
- MILLER, H., & BIERI, J. Cognitive complexity as a function of the significance of the stimulus objects being judged. *Psychological Reports*, 1965, **16**, 1203-1204.
- POSTMAN, L. The experimental analysis of motivational factors in perception. In *Current theory and research in motivation*. Lincoln: University of Nebraska Press, 1953. Pp. 59-108.
- TRIPODI, T., & BIERI, J. Cognitive complexity as a function of own and provided constructs. *Psychological Reports*, 1963, **13**, 26.

(Received July 6, 1966)

SOME SOCIAL DETERMINANTS OF SELF-MONITORING REINFORCEMENT SYSTEMS¹

ALBERT BANDURA, JOAN E. GRUSEC, and FRANCES L. MENLOWE

Stanford University

The present experiment was designed to establish the social conditions under which children who observe high standards of self-reward come through the adoption of such contingencies assuming negative self-evaluative consequences. Children were exposed to an adult role model who achieved superior performance and adopted a high standard of self-reward. The main variable subject to study was the degree of self-reward imposed by the model, and the extent of the model's high achievement behavior, and the presence or absence of a peer model who modeled a low standard of self-reward. Various positive reinforcements of model's behavior produced a severe pattern of demands for self-reward, whereas low model performance and exposure to modeling processes had no effect, relative to adult modeling cues. The combined influence of low achievement, elevated position on treatment, and the absence of modeling processes, which produced the most stringent pattern of self-reward.

Successful socialization and personality development require the gradual substitution of self-monitoring reinforcement systems for externally imposed sanctions and demands. In order to achieve such self-regulation, persons must adopt certain standards of behavior and self-administer rewards and punishments depending upon whether their performances fall short of, match, or exceed their self-prescribed demands.

Findings of research conducted within a social learning framework (Bandura & Kupers, 1964; Bandura & Whalen, 1966; Marston,² 1965; Mischel & Liebert, 1966) demonstrate that self-reinforcement systems can be transmitted on the basis of observational learning through exposure to the standard-setting behavior and self-reinforcing patterns exhibited by adult and peer models. There is some recent evidence (Bandura & Whalen, 1966), however, that modeling variables alone are not sufficient for imparting high standards of self-reward as exemplified by highly competent persons. This finding is not entirely unexpected since an individual who adopts stringent criteria of self-reward will more often judge his performances to

be inadequate than if he adheres to lower standards of achievement. High self-imposed demands, therefore, inevitably result in more frequent self-denial of available rewards and negative self-evaluative consequences (Bandura & Kupers, 1964). For this reason, persons are generally inclined to reject the self-reinforcement contingencies displayed by superior models. On the other hand, it is evident from both informal observation and the variable responsivity of subjects in laboratory studies (Bandura & Whalen, 1966) that many persons do, in fact, emulate exacting self-reward standards even though such behavior increases the probability of unfavorable self-generated consequences.

The present experiment was primarily designed to investigate some of the social conditions which might lead children to adopt stringent self-reward patterns of behavior exhibited by superior models.

There are three classes of variables that are likely to serve as important determinants of modeling processes. The first of these is the nature of the relationship between the model and the observer. Among the numerous relationship factors that have received attention, the nurturant or rewarding quality of the model, which tends to increase interpersonal attraction, has been shown to be influential in facilitating identificatory outcomes (Bandura & Huston, 1961; Henker,

¹ This research was supported by Public Health Research Grant M-5162 from the National Institute of Mental Health.

The authors are grateful to Alec Smith, director of the Palo Alto Recreation Department, for his assistance in arranging the research facilities.

1964; Mischel & Grusec, 1966; Mussen & Parker, 1965).

The extent to which a model's behavior will be reproduced by others may also be greatly influenced by the reinforcing consequences associated with the critical response patterns. There is little question but that children would conform to the standards of achievement exemplified by a model if extrinsic incentives of sufficient positive valence were made contingent upon matching behavior. Of greater social and theoretical significance, however, is the spontaneous emulation of models that results from witnessing reinforcing consequences, usually in the form of social approval and public recognition, accruing to the model (Bandura, 1965b, 1967). Therefore, the role of vicarious reinforcement in the social transmission of self-reward contingencies was the second variable selected for study.

In naturalistic situations individuals are typically confronted with a multiplicity of modeling influences, many of which operate in opposing directions. Theoretical speculations about the effect of multiple modeling on social learning generally give considerable emphasis to the conflicting identifications occurring in relation to adult and peer models. In the case of achievement standards and self-reinforcing patterns of behavior, there are several factors that might predispose children toward peer modeling when they are confronted with a conflict in standards between adults and peer members. If adult achievements and criteria for self-reward are relatively high, as is usually the case, then children will be reluctant to emulate such high aspirations because to do so would result in frequent negative self-reinforcement of their performances. In addition, according to social comparison theory, adults are likely to be viewed by children as too divergent in ability to serve as meaningful reference models (Bandura & Whalen, 1966). However, it is possible that conditions serving to reduce receptivity of adult modeling cues might be effectively counteracted by the simultaneous operation of opposing influences arising from strong affective ties to the model, and from vicarious positive reinforcement of high standard-setting behavior.

In the experiment reported in this paper, children were exposed to an adult model who performed a task at a consistently superior level relative to that of the children and adopted a high criterion of self-reward. Half the subjects experienced a prior rewarding interaction with the model, whereas for a second group of children the same model behaved in a nonnurturant manner. With half the subjects in each of the two levels of nurturance, the model was praised for adhering to stringent standards of achievement, but with the remaining children the model received no social recognition for high standard-setting behavior. In addition, half the children in each subgroup also observed a peer model who displayed a low standard of self-reward in order to determine the effects of simultaneous exposure to conflicting modeling cues.

It was predicted that model nurturance and vicarious positive reinforcement would augment, and peer modeling would reduce, the emulation of stringent self-reward patterns of behavior. Moreover, it was hypothesized that the combined influence of nurturance, vicarious reinforcement, and the absence of conflicting modeling cues would produce the highest level of identificatory behavior.

METHOD

Subjects

The sample, drawn from four elementary schools participating in the Palo Alto summer recreation program, contained 64 boys and 64 girls ranging in age from 7 to 11 years.

The children were divided into groups of male and female subjects and randomly assigned to the various experimental subgroups. A $2 \times 2 \times 2 \times 2$ factorial design was used, with 2 levels of nurturance, 2 conditions of modeling cues, 2 conditions of vicarious reinforcement of the self-rewarding behavior displayed by the adult model, and the sex differentiation of the subjects.

A male and a female adult and four children between the ages of 8 and 10 years served in the role of models.² All subjects in the experiment were exposed to same-sex models.

Nurturant Treatment

The children were brought individually to a mobile laboratory, ostensibly to test some game

² The authors wish to thank Mary Bierman Harris, Ronald Kaufman, Sherrie and Robert Cutler, Patti Porter, and Jonathan Miller for their faithful modeling services.

equipment that had been designed for use by both children and adults. After introducing the child to the adult model, who presumably was waiting his turn to play the game, the experimenter excused herself in order to prepare the equipment.

With half the children, who were assigned to the low-nurturance condition, the model announced that he would read his newspaper while they waited, and he supposed that the toys on the table were provided for the children to play with. The model then proceeded to read his newspaper for the entire 15-minute period and refrained from any social interaction with the child.

By contrast, in the high-nurturance condition, after the experimenter departed, the model remarked that he had planned to read his newspaper, but since the child was there they might as well play together. The model then obtained some additional toys from an adjacent room and played actively with the subject throughout the session. During the high level of social interaction, the model responded in a consistently warm, friendly, and generously rewarding manner.

Modeling of Conflicting Self-Reward Contingencies

Immediately following the nurturance manipulation, the experimenter returned and escorted both the adult model and the subject to another room in the mobile laboratory that contained the bowling apparatus utilized for modeling and measuring self-reward responses.

The apparatus, which has been described elsewhere in detail (Bandura & Whalen, 1966), consisted of a miniature bowling alley bounded at the far end by a vertical shield. Eight jewel lights, labeled with numbers ranging from 10 to 80, were mounted in three staggered rows on the front shield. The subjects were informed that whenever a bowling ball hit a target (purportedly behind the shield) the corresponding score light would flash on. In fact, the scores were preset from a control panel located in the adjoining observation room, so that the models performed identically in each condition, and all subjects received the same pattern of scores.

Before commencing the modeling trials, the experimenter called the participants' attention to a large bowl of plastic tokens near the starting point of the alley. The participants were informed that they were free to treat themselves to tokens whenever they felt that they had performed well and that, at the conclusion of the game, the tokens would be exchanged for prizes: the more tokens they obtained, the more valuable the prizes. The models were each handed a bank in which to deposit their chips, and again given permissive instructions to reward themselves with the token currency whenever they judged their bowling performance to be satisfactory.

Half the children in both the high- and low-nurturance conditions were exposed simultaneously to an adult and a peer model, who displayed conflicting standards for self-reward. The two models

alternated in blocks of five trials, for a total of 20 games each, while the subject, who presumably was waiting for his adult partner to arrive, observed their self-rewarding patterns of behavior.

The adult model achieved superior performances ranging from 50 to 80 points and rewarded himself with tokens and positive self-evaluative comments only when he obtained or exceeded the relatively high score of 60 points. On the occasions when the model attained his self-imposed standard, he commented approvingly, "That's a good score, I deserve a chip for that," and rewarded himself with one token. Following performances above the adopted criterion level, the model engaged in more generous self-reward: "That was really a good score. I deserve two chips for that." On the other hand, after trials in which he failed to meet the adopted criterion of 60, he denied himself tokens and remarked self-critically, "A score like that does not deserve a chip."

By contrast, the peer model obtained scores ranging from 10 to 40 points and adopted the relatively low self-reward criterion of 20 points. Except for the lower standard, the peer model exhibited the same pattern, magnitude, and frequency of self-reinforcement as his adult counterpart. On trials in which the peer obtained or exceeded a score of 20, he rewarded himself with tokens and self-approving comments, whereas on trials in which he failed to meet the self-imposed standard he took no tokens and criticized himself.

Children who were assigned to the peer-absent condition observed the adult model exhibit the same stringent pattern of self-reward, but they were not exposed to the low standards set by the peer model.

Vicarious Reinforcement of Stringent Self-Reward Demands

At the conclusion of the modeling trials, with half the children within each subgroup, the experimenter praised the adult model in the subject's presence for adopting and adhering to high standards of performance. The social reinforcement was primarily in the form of approving comments about the fact that the adult model was a person who set himself high standards of achievement and thought well of himself only when he had done an excellent job. In conditions involving a peer model, the adult was commended after the peer had been thanked for his participation and had departed, in order to minimize any implied negative sanction of the peer's more lenient self-rewarding tendencies.

For the remaining subjects within each experimental subgroup, the adult model was merely thanked for his assistance, but received no social recognition for his high standard-setting behavior.

Measurement of Self-Rewarding Responses

In treatments that included the peer model, after both models had departed, the experimenter explained to the subject that his adult partner was

apparently delayed for his scheduled session and, therefore, the child could play the game alone. In the peer-absent conditions, the experimenter told each child that it was now his turn to play with the bowling apparatus. The experimenter then replenished the token supply, handed the child his bank, described the token exchange, and repeated the permissive instructions for self-reward.

In order to remove any possible external influences on subjects' self-reinforcing responses, the postexposure test was conducted with each child alone in the room. The experimenter explained that she had some work to complete in another room of the laboratory, and that the child might continue playing the game until she returned.

After the experimenter left, the children performed 36 trials and obtained the same pattern of scores, ranging from 10 to 60 points, as determined by a prearranged program. It might be expected that persons would readily adopt and adhere to high standards of self-reward under conditions where their performances are sufficiently high to ensure frequent opportunities for positive self-reinforcement. The present experiment, however, was primarily concerned with testing the efficacy of social learning variables in fostering the adoption of high standards under circumstances where modeling produces sparing self-reinforcement, as is typically the case in naturalistic situations. Therefore, the children's scores were considerably lower than those obtained by the adult model, and their best performances matched the model's minimum criterion of 60 points on only six trials.

In the adjoining observation room, one observer recorded the performances for which the children rewarded themselves with tokens and the number of tokens taken on each self-reinforced trial. In order to determine scorer reliability, the self-rewarding responses of 12 subjects were recorded independently by a second observer. The two raters achieved virtually perfect agreement (99.4%) in scoring both the incidence and the magnitude of self-reinforcement.

RESULTS

Table 1 presents the mean percentage of trials in which children in the various experimental groups rewarded themselves for performances below the 60 points adopted by the adult model as the minimum criterion of self-reward. These data provide the best overall index of the degree to which exposure to stringent self-reinforcement contingencies exhibited by adults under different social conditions influenced children's standard-setting and self-reinforcing behavior.

Analysis of variance performed on these scores disclosed that highly significant effects were produced by the behavior modeled by the peer ($F = 37.64$, $p < .001$), by posi-

TABLE 1
MEAN PERCENTAGE OF TRIALS IN WHICH SUBJECTS REWARDED THEMSELVES FOR PERFORMANCES BELOW THE MODEL'S MINIMUM CRITERION OF SELF-REWARD

Subgroup	Social approval		No approval	
	Rewarding interaction	Nonrewarding interaction	Rewarding interaction	Nonrewarding interaction
Peer model				
Boys	57	39	71	66
Girls	41	45	60	62
Total	49	42	66	64
No peer model				
Boys	25	10	45	37
Girls	35	12	40	31
Total	30	11	43	34

tive reinforcement of the adult model ($F = 20.32$, $p < .001$), and, to a lesser degree, by the nurturant interaction between the adult model and the children ($F = 4.64$, $p < .05$). In accord with prediction, children who were exposed to conflicting modeling cues were more inclined to reward themselves for low achievements (55% of the trials below 60) than children who had observed only the adult models adhere to high standards of self-reinforcement (30%). Similarly, children imposed more severe criteria of self-reward on themselves when the adult model received social recognition for his high standard-setting behavior (33%) than when the model's stringent achievement demands went unrewarded (52%). However, contrary to expectation, subjects who had experienced a highly nurturant interaction with the adult model were somewhat more inclined to accept the lower standard set by the peer (47%) than if the adult model were less beneficent (38%). The analysis, however, yielded no sex differences or significant interaction effects.

As shown in Table 2, except for a few interesting variations, the variables discussed above influenced children's self-reinforcing responses in essentially the same manner at each specific performance level. Vicarious reinforcement decreased, and peer modeling cues increased, the incidence of self-reinforcing responses, with the effects being particularly powerful at lower and intermediate

TABLE 2
F VALUES OBTAINED AT EACH PERFORMANCE LEVEL FOR THE
DIFFERENT EXPERIMENTAL VARIABLES

Experimental variables	Performance level					
	10	20	30	40	50	60
Vicarious reinforcement	5.20**	13.17***	20.07***	14.60***	8.46***	3.91**
Peer model	<1	26.00***	54.84***	50.00***	8.09***	3.91**
Nurturance	<1	2.89*	5.65**	3.21*	1.68	<1
Sex	<1	<1	<1	2.64	<1	<1

* $p < .10$.

** $p < .05$.

*** $p < .01$.

**** $p < .001$.

levels of achievement. It is interesting to note, however, that exposure to the peer model did not increase the children's frequency of self-reward for performances that fell below the minimum standard adopted by the peer (Table 2). It is also noteworthy that model nurturance increased subjects' tendency to reward themselves at the intermediate range of achievement, but it did not exert a significant effect at either low or high performance levels.

Magnitude of Self-Reward

The mean number of rewards that the children administered to themselves on each self-reinforced trial as a function of experimental conditions is shown in Table 3.

The results of the analysis of variance performed on these means reveal that the experimental variables affected the magnitude as well as the frequency of self-reinforcement.

TABLE 3

MEAN NUMBER OF TOKENS TAKEN BY CHILDREN IN
THE DIFFERENT EXPERIMENTAL CONDITIONS
ON EACH SELF-REINFORCED TRIAL

Subgroup	Social approval		No social approval	
	Rewarding interaction	Nonrewarding interaction	Rewarding interaction	Nonrewarding interaction
Peer model				
Boys	1.61	1.73	1.63	1.89
Girls	1.58	1.44	2.35	1.57
Total	1.59	1.58	1.99	1.73
No peer model				
Boys	1.29	1.03	1.76	1.40
Girls	1.33	1.30	1.32	1.61
Total	1.31	1.17	1.54	1.50

Children who witnessed peers self-administer reinforcers for minimal performances rewarded themselves more generously than subjects whose exposure was limited to the high contingencies exemplified by adult models ($F = 17.59$, $p < .001$). On the other hand, commending a model for adhering to lofty standards resulted in more sparing self-reward by observers ($F = 10.87$, $p < .01$). In addition to these main effects, the analysis disclosed a significant Peer \times Sex \times Nurturance interaction effect ($F = 10.78$, $p < .01$). For girls, nurturance increased generosity of self-reward in the peer condition, but produced less liberal self-reinforcement when there was no exposure to the conflicting peer standards; the opposite pattern of relationships was obtained for boys.

Since some of the children never rewarded themselves after obtaining scores below 60, the number of cases in each cell varied somewhat from group to group at different score levels. Consequently, separate one-way analyses of variance were computed for evaluating the influence of vicarious reinforcement, peer modeling cues, nurturance, and sex of subjects on magnitude of self-reward at each specific performance level. Although all groups rewarded themselves sparingly for low accomplishments, and did not differ in this respect, children who had observed the peer model subsequently treated themselves more generously whenever they received scores in the 50s ($F = 19.12$, $p < .001$) or 60s ($F = 7.53$, $p < .01$). Children in the vicarious reinforcement condition, on the other hand, attached less value to their achievements at the 40

($F_1 = 3.63$, $p < .10$), 50 ($F = 7.30$, $p < .01$), and 60 ($F = 4.86$, $p < .05$) performance range.

DISCUSSION

This experiment provides some indication of the relative influence of model nurturance, vicarious positive reinforcement, and peer modeling cues on emulation of stringent self-reward patterns of behavior.

In most theories of identification (Bronfenbrenner, 1960; Sears, Rau, & Alpert, 1965) interpersonal relationship variables receive considerable attention, but the influential role of response consequences to the model in determining identificatory outcomes has been virtually ignored. The findings of this study further demonstrate that social reinforcement of a model's responses can have a powerful effect on observers' spontaneous reproduction of matching behavior. Children who had observed the adult model adopt high achievement standards for self-reward, and receive social recognition for adhering to such exacting norms, subsequently imposed on themselves higher performance demands than children who had witnessed the model portray the same pattern of self-reinforcement without any socially rewarding consequences.

Previous studies (Bandura, 1965a; Walters & Parke, 1964; Walters, Parke, & Cane, 1965) comparing the effects of exposure to rewarding, punishing, or no consequences to a model for engaging in socially censured behavior provided no clear evidence for the occurrence of positive vicarious reinforcement. The omission of adverse consequences following displays of reprehensible behavior can function to reduce observers' inhibitions, with the result that nonoccurrence of anticipated punishing reactions typically augments matching behavior to the same degree as witnessed rewarding outcomes. Findings of the present experiment, involving socially approved behavior in which inhibitory processes are likely to play a minor role, show that social rewards dispensed to a model produce a higher incidence of matching behavior than exposure to the same modeling cues without any consequences accruing to the model.

Model nurturance not only exerted a some-

what weaker influence on children's self-rewarding behavior compared to the other variables, but also, contrary to expectation, children who had experienced a highly nurturant interaction with the adult model were more inclined to accept the low performance standards set by the peer than if the adult were less beneficent. It thus appears that in the case of evaluative standards that govern the incidence of self-rewarding behavior, high nurturance is conducive to ready self-gratification rather than to emulation of stringent achievement demands self-imposed by the rewarding adult.

The latter findings, and results of other studies in which the rewarding quality of the model is varied experimentally, suggest that the developmental or anacletic theory of identification, which assumes that model nurturance enhances identification, may be valid only under certain limiting conditions. A study by Bandura and Huston (1961) disclosed that, although a model's rewarding quality facilitated reproduction of verbal and stylistic responses, children readily adopted aggressive responses regardless of the degree of the model's nurturance. Similarly, Mischel and Grusec (1966) found that a prior nurturant interaction with the model enhanced children's spontaneous imitation of socially neutral behaviors, but it did not increase their willingness to perform matching responses that possessed aversive properties. As noted earlier, adoption of high self-evaluative standards likewise involves some aversive effects. The implication of these overall results is that model nurturance may produce specific rather than generalized modeling effects, and that it may differentially influence the reproduction of neutral and negatively valenced classes of behavior.

The experiment also revealed some of the conditions under which exposure to the performance norms of peer models has a significant effect upon emulation of adult standards. Children who simultaneously observed an adult set high criteria of achievement for himself and a peer adopt low norms subsequently imposed lower self-reward contingencies on themselves and rewarded themselves more generously than children who were exposed only to the behavior of the adult

model. However, the influence of the peer's lenient pattern of self-reward was effectively counteracted by social reinforcement of the adult's high standard-setting behavior. This neutralization of influences is shown most clearly in the finding that the condition including the joint effect of peer modeling cues and vicarious reinforcement yielded essentially identical rates of self-reinforcement as the treatment in which the powerful opposing influences of peer cues and adult approval were both absent.

The identification process is particularly complicated under naturalistic conditions in which children are exposed to adult standards as well as multiple peer models who display conflicting patterns of behavior. It would, therefore, be of considerable interest to isolate the social variables determining whether children will select peer models who reinforce the adult standards or those who furnish opposing influences.

The most austere pattern of self-reward was displayed by children who had experienced a relatively nonnurturant relationship with the model, who had no exposure to conflicting standards of peer models, and who witnessed the adult's high standard-setting behavior socially reinforced. Approximately half the children in this condition never rewarded themselves for performances that fell below the adult's criterion, while the remainder rarely considered scores below 50 deserving of self-reward. The adoption and continued adherence to unrealistically high self-evaluative standards is particularly striking considering that the self-imposition of rigorous performance demands occurred in the absence of any social surveillance, under high permissiveness for self-gratification, and the emulative behavior resulted in self-critical reactions and considerable self-denial of freely available rewards. These findings provide further evidence that inhibitions and strong self-controlling responses may be acquired through observational learning without the mediation of direct positive or negative reinforcement.

REFERENCES

- BANDURA, A. Influence of models' reinforcement contingencies on the acquisition of imitative responses. *Journal of Personality and Social Psychology*, 1965, 1, 589-595. (a)
- BANDURA, A. Vicarious processes: A case of no-trial learning. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 2. New York: Academic Press, 1965. Pp. 1-55. (b)
- BANDURA, A. Social-learning theory of identificatory processes. In D. A. Goslin & D. C. Glass (Eds.), *Handbook of socialization theory and research*. Chicago: Rand McNally, 1967, in press.
- BANDURA, A., & HUSTON, A. C. Identification as a process of incidental learning. *Journal of Abnormal and Social Psychology*, 1961, 63, 311-318.
- BANDURA, A., & KUPERS, C. J. The transmission of patterns of self-reinforcement through modeling. *Journal of Abnormal and Social Psychology*, 1964, 69, 1-9.
- BANDURA, A., & WHALEN, C. K. The influence of antecedent reinforcement and divergent modeling cues on patterns of self-reward. *Journal of Personality and Social Psychology*, 1966, 3, 373-382.
- BRONFENBRENNER, U. Freudian theories of identification and their derivatives. *Child Development*, 1960, 31, 15-40.
- HENKER, B. A. The effect of adult model relationships on children's play and task imitation. *Dissemination Abstracts*, 1964, 24, 4797.
- MARSTON, A. R. Imitation, self-reinforcement, and reinforcement of another person. *Journal of Personality and Social Psychology*, 1965, 2, 225-261.
- MISCHEL, W., & GRUSEC, J. Determinants of the rehearsal and transmission of neutral and aversive behavior. *Journal of Personality and Social Psychology*, 1966, 3, 197-205.
- MISCHEL, W., & LIEBERT, R. M. Effects of discrepancies between observed and imposed reward criteria on their acquisition and transmission. *Journal of Personality and Social Psychology*, 1966, 3, 45-53.
- MUSSEN, P. H., & PARKER, A. L. Mother nurturance and girls' incidental imitative learning. *Journal of Personality and Social Psychology*, 1965, 2, 94-97.
- SEARS, R. R., RAU, L., & ALPERT, R. *Identification and child rearing*. Stanford: Stanford University Press, 1965.
- WALTERS, R. H., & PARKE, R. D. Influence of response consequences to a social model on resistance to deviation. *Journal of Experimental Child Psychology*, 1964, 1, 269-280.
- WALTERS, R. H., PARKE, R. D., & CANE, V. A. Timing of punishment and the observation of consequences to others as determinants of response inhibition. *Journal of Experimental Child Psychology*, 1965, 2, 10-30.

(Received July 25, 1966)

BRIEF ARTICLES

INFORMATION SEARCH AS A FUNCTION OF CONCEPTUAL STRUCTURE IN A COMPLEX PROBLEM-SOLVING TASK¹

MARVIN KARLINS AND HELMUT LAMM

Princeton University

Individuals varying in their level of integrative complexity (Schroder, Driver, & Streufert) requested information about a novel environment for use in solving a complex problem. The findings of the investigation support the contention that when Ss varying in integrative complexity are allowed to learn a task environment by self-instigated information search they approach it differently. Ss who are integratively complex are more active in this type of environment (ask more questions) than their integratively simple counterparts. The experimental findings are discussed in relation to creativity research, and other studies examining the amount of information search as a function of integrative complexity. It is suggested that the reason some studies do not find more active information search on the part of integratively complex Ss is due to the kind of task Ss are required to perform.

In the past decade personality theorists have become increasingly concerned with the ways individuals explore the dimensions of a formerly unknown environment or problem. This concern with an individual's pattern of *information search* has led to a new perspective in studying problem solving. Whereas in the past, concern was centered on the end product of problem solving—the answer—emphasis is now being placed on the earlier phases of problem solving: the ways individuals search for information in solving the problem. The question becomes not how much information is known, but how that information is acquired in dealing with the unknown. This new orientation is reflected in the "discovery" or "inductive" learning methods utilized in education (Anderson & Moore, 1960; Suchman, 1961).

One theory of personality concerned with the ways individuals search for information is the structural complexity approach of Schroder and his associates (Harvey, Hunt, & Schroder, 1961; Schroder, Driver, & Streufert, 1966). In this model, any psychological situation may be simply described as a range of stimuli which is "read" and transformed via mediating processes into a set of behavioral outputs or responses. There are several characteristics that describe the functioning of these mediating processes, the most important of which is the information processing of the system. Over a given range of stimuli,

information-processing ability varies among individuals and is measured in terms of its *integrative complexity*. Higher integrative complexity refers to a greater number of perceptual categories for receiving information about the world and more conceptual or combinatory rules for organizing such units of information. Integrative complexity is described as varying along a continuum that represents gradations in information-processing ability. It is measured by the Paragraph Completion Inventory (Schroder & Streufert, 1963; Schroder et al., 1966, Ch. 12), a semiprojective test that requires subjects to write three-sentence responses to stimuli that create discrepancy and uncertainty.

The Schroderian information-processing conception can be utilized to make predictions about the way an individual will search for information, based on his level of integrative complexity. In *complex*² problem-solving tasks it is expected that integratively complex individuals will be able to bring a greater number of dimensions to

² The necessity of providing a *complex* problem-solving task is based on theoretical considerations. It is a major contention of structural complexity thinking that the level of environmental complexity influences the degree to which individual *performance* differences in the acquisition of information can be manifested in a given problem-solving situation. It is hypothesized that

... the general relationship between environmental complexity and the amount of differentiation and integration involved in information processing is curvilinear for persons at any level of structural complexity [Schroder et al., 1966].

¹ This research was supported by funds from the National Institute of Mental Health and the Peace Corps. We thank Harold Schroder and Bertram Koslin for their helpful suggestions.

bear on a given problem and generate more perspectives that link specific dimensions with task considerations. It is predicted that integratively complex individuals, in comparison to integratively simple individuals, will be *more active in their information search* (ask more questions) in dealing with situations that require the generation of perspectives.

This hypothesis has received very little experimental attention, and the findings of the few studies that have investigated it are inconclusive. The three major studies dealing with this question have all utilized the tactical war game of Streufert, Clardy, Driver, Karlins, Schroder, and Suedfeld (1965), one of which used the "fixed input program" modification of Karlins, Schroder, and Streufert (1965). In these investigations information search was defined in various ways, the most common measure being the number of self-initiated information requests made by the subjects. When these requests concerned past or ongoing events, they were considered to be *monitoring* questions. When the requests dealt with situations or locations where subjects had not been involved previously (Stager, 1967; Suedfeld & Streufert, 1966) or future decisions which grew out of previous information (Streufert, Suedfeld, & Driver, 1965), they were labeled *novel* or *integrative information questions*.

When the total number of self-initiated information requests were considered without regard to their content (e.g., monitoring or novel), then the Stager (1967) and Streufert, Suedfeld, and Driver (1965) studies found no differences in information search between integratively complex and simple subjects. Suedfeld and Streufert (1966) found that integratively simple subjects made *more* informational requests in periods of the tactical situation when stress (information load) was low and moderate, but not when it was high. All the studies, however, found that integratively complex subjects were superior to integratively simple subjects in their information search when, instead of considering the *number* of information requests, the *novelty* or *integrative information* of the questions was analyzed. From these studies it would seem that mere *amount* of information search (more active information search) did not differentiate subjects varying in integrative complexity. Rather, the *kind* of information requested became the critical distinguishing variable.

These findings do not lend support to the present hypothesis which considers the volume of information search regardless of the content of that search. It might be that the tactical war game is not the most suitable task for

testing a volume factor in information search. In the war game one type of information search was really more information *re-search* (e.g., "monitoring questions" were often no more than requests for progress reports about actions already taken). This type of information search does *not* require the subject to generate perspectives in dealing with his problem. In this light the findings of the three studies make sense. One would *not* predict information-search differences between subjects varying in structural complexity where they do not have to generate questions that require linkages between perspectives and the task situation.

A new problem-solving situation, the Community Development Task, has been developed to retest the volume factor in information search in a problem-solving framework where each type of question asked would be analogous to the type considered "novel" or "integrative information search," in the terminology of the war-game investigators. It is hypothesized that integratively complex subjects, in comparison with integratively simple subjects, will undertake more extensive information search as measured by the total number of questions asked during 135 minutes of information acquisition.

METHOD

Subjects

Sixty subjects, matched on intelligence and varying in conceptual level (30 integratively simple, 30 integratively complex), participated in the experiment. Intelligence was assessed by the Wonderlic Personnel Test, an intelligence test that correlates .87 with the Otis Quick-Scoring Mental Ability Test. Integrative complexity was measured by the Paragraph Completion Inventory (Schroder et al., 1966). All subjects were Peace Corps volunteers, the majority being college graduates in their early 20s.

Experimental Task

In the Community Development Task each subject was asked to imagine that he was a Peace Corps volunteer assigned to build, in a cooperative effort with the local natives, a hospital on an isolated South Seas island. He was informed that a similar task had been attempted by missionaries 3 years before, but that these individuals were unable to enlist native aid, and the project failed. The importance of getting the hospital built was emphasized to each subject: venereal disease was spreading rapidly on the island—killing the natives and infecting traders in the area. Utilizing his information-search prerogatives, each subject was expected to learn about the culture of the islanders he would be living with, and in such a way present a tentative solution to the problem, that is, his estimation of the *best possible way* to get native approval and cooperation in building the hospital.

The subject's job was further complicated by the fact that he began the task with no knowledge of the island or its inhabitants. The only way he could gain information he felt relevant to his task was through active information search: written questions submitted to the experimental staff. The answers to his questions (one or two typewritten sentences) were returned to him on the back of coded IBM cards.³ The subject understood that the only way he could obtain answer cards was by asking questions about his task: the number and type of answer cards the subject could receive were entirely determined by the number and type of questions he asked.

There were 1,250 answer cards, prepared in advance of the experiment, which the subject could receive in three 45-minute periods allotted to information acquisition about the hospital-construction problem. These answer cards contained information about all phases of the island culture. The information on the cards was based on Malinowski's (1922) work with the Trobriand islanders and the cultural categories provided by the Human Relations Area Files (Murdock, Ford, Hudson, Kennedy, Simmons, & Whiting, 1950).

RESULTS AND DISCUSSION

The finding, bearing on the hypothesis that integratively complex individuals (in comparison to integratively simple subjects) would be more active in their information search, as reflected in the number of questions asked in a complex problem-solving task, is presented in Table 1. The results are in support of the hypothesis at the .025 level of significance ($t = 2.33$).

These findings lend credence to the notion that when subjects varying in integrative complexity are allowed to learn a task by active manipulation of the environment, they approach that environment differently. Subjects who are integratively complex are more active in this interdependent environment, asking more questions concerning their problem. In the present problem, with many degrees of freedom and a good deal of complexity, it is argued that the integratively complex individual will ask more questions because he has more perspectives to assist him in generating questions. In tasks that do not require this form of information acquisition, no performance differences by individuals varying in integrative complexity would be expected.

Finally, the results of the present investigation have some interesting ramifications in the field of creativity research. Some creativity theo-

³ In all phases of the experiment the subject was physically separated from the experimenter by a one-way mirror.

TABLE 1
COMPARISON OF NUMBER OF QUESTIONS ASKED BY
SUBJECTS LOW AND HIGH IN INTEGRATIVE
COMPLEXITY

Low		High		t
\bar{X}	SD	\bar{X}	SD	
56.60	20.49	69.26	20.78	2.33*

Note.— $N = 30$ for both low and high subjects.

* $p < .025$, 1-tailed.

rists define creativity as proficiency in generating various products, such as verbal associates (Mednick, 1962) and uses for things (Guilford, 1959; Wallach & Kogan, 1965). In a way, the ability to produce a number of questions about an unknown environment might itself be considered an aspect of creativity. Certainly the ability to formulate questions about a problem is an important first step in the problem-solving process.

REFERENCES

- ANDERSON, A. R., & MOORE, O. K. Autotelic folk-models. *Sociological Quarterly*, 1960, 1, 203-216.
- GUILFORD, J. P. Three faces of intellect. *American Psychologist*, 1959, 14, 469-479.
- HARVEY, O. J., HUNT, D. E., & SCHRODER, H. M. *Conceptual systems and personality organization*. New York: Wiley, 1961.
- KARLINS, M., SCHRODER, H. M., & STREUFERT, S. *A fixed input program for a tactical war game*. Washington, D. C.: American Documentation Institute, Auxiliary Publications Project, Photoduplication Service, Library of Congress, 1965, No. 8621.
- MALINOWSKI, B. *Argonauts of the western Pacific*. New York: Dutton, 1922.
- MEDNICK, S. A. The associative basis of the creative process. *Psychological Review*, 1962, 69, 220-232.
- MURDOCK, G. P., FORD, C. S., HUDSON, A. E., KENNEDY, R., SIMMONS, L. W., & WHITING, J. W. *Outline of cultural materials*. (3rd ed.) New Haven, Conn.: Human Relations Area Files, Inc., 1950.
- SCHRODER, H. M., DRIVER, M. J., & STREUFERT, S. *Human information processing: Individuals and groups in complex social situations*. New York: Holt, Rinehart & Winston, 1966.
- SCHRODER, H. M., & STREUFERT, S. The measurement of four systems of personality structure varying in level of abstractness: Sentence completion method. Technical Report No. 11, Project NR 171-055, 1963, Princeton University, Contract NQNR 1858(12), Office of Naval Research.
- STAGER, P. Conceptual level as a composition variable in small-group decision making. *Journal of Personality and Social Psychology*, 1967, 5, 175-185.

- STREUFERT, S., CLARDY, M. A., DRIVER, M. J., KARLINS, M., SCHRODER, H. M., & SUEDELD, P. A tactical game for the analysis of complex decision making in individuals and groups. *Psychological Reports*, 1965, 17, 723-729.
- STREUFERT, S., SUEDELD, P., & DRIVER, M. J. Conceptual complexity, information search, and information utilization. *Journal of Personality and Social Psychology*, 1965, 2, 736-740.
- SUCHMAN, R. J. Inquiry training: Building skills for autonomous discovery. *Merrill-Palmer Quarterly*, 1961, 7, 147-169.
- SUEDELD, P., & STREUFERT, S. Information search as a function of conceptual and environmental complexity. *Psychonomic Science*, 1966, 4, 351-352.
- WALLACE, M. A., & KOGAN, N. *Modes of thinking in young children*. New York: Holt, 1965.

(Early publication received November 16, 1966)

Journal of Personality and Social Psychology
1967, Vol. 5, No. 4, 459-463

EFFECT OF SIMILARITY AND FORTUNE OF THE OTHER ON ATTRACTION¹

JERALD M. JELLISON AND JUDSON MILLS

University of Missouri

An experiment was conducted to investigate how the tendency to like similar persons more than dissimilar persons is affected by the fortune of the other person. College women evaluated another woman from the other's answers to a questionnaire they themselves had previously completed. Some Ss received a questionnaire with answers exactly the same as their own; others, a questionnaire with answers opposite to their own. Some were informed the woman had won a free trip to the Paris fashion showings; others, that nothing unusual had happened to the woman. Contrary to prediction, the tendency to like a similar other more than a dissimilar other was greater when the other had good fortune than when the other had average fortune.

A number of experiments (Byrne, 1961; Byrne & Nelson, 1965; Jones & Daugherty, 1959; Smith, 1957) have shown that people tend to like another person who they think has attitudes similar to their own more than someone with dissimilar attitudes. The present study is concerned with the effect of the fortune of the other person on the tendency to like similar others more than dissimilar others. It was hypothesized that the tendency would be smaller when the other has some very good fortune than when the other has only average fortune.

The stimulus for the hypothesis came from Heider's (1958) discussion of the reaction to the lot of the other person, in particular some of his comments concerning envy.

If p considers himself in the same class as o then there is a strong tendency also to consider that their fortunes should be the same. . . . Thus, should o have a stroke of luck, then p may feel unhappy; he wants to experience the positive x also, and if possible will try to share it with o ; p may think that o ought not to have the x ; he may think that

the x is not so good after all, that it will not make o happy; because this negative experience of p may be connected with o as its source, a negative sentiment of p toward o may arise. . . . This tendency toward the equalization of lot will be recognized as the common form and source of envy. As was mentioned above, the tendency occurs where p feels bound up with o in some way, similar to o [Heider, 1958, p. 287].

If people are more likely to envy a person with good fortune when they think he is similar to themselves than when they think he is dissimilar, and if envy creates dislike for the other person, then the tendency to like similar others more than dissimilar others should be smaller when the other has very good fortune than when the other has only average fortune.

METHOD

One week after undergraduate college women had filled out a self-descriptive questionnaire, they were asked to rate the personal qualities of another person from that person's answers to the same questionnaire. They were given fake questionnaires designed to create four experimental conditions: high similarity-good fortune, low similarity-good fortune, high similarity-average fortune, and low similarity-average fortune. The dependent variable, attraction,

¹ This study was conducted while the first author held a United States Public Health Service predoctoral fellowship.

was measured by means of ratings of how much the other person was liked and how well favorable and unfavorable personal adjectives applied to the other. The subjects were 86 freshman and sophomore women in four beginning psychology classes at Stephens College.² The experiment was conducted during the regular class sessions. Within each of the classes, subjects were randomly assigned to the four experimental conditions, with the provision that subjects sitting next to one another be in different conditions.

On his first appearance before a class, the experimenter administered a questionnaire, entitled the Personal Inventory, under the guise of studying the personal preferences of college students. The Personal Inventory contained 50 items with two alternatives each, selected from the Edwards (1953) Personal Preference Schedule (EPPS). The final item on the Personal Inventory was an open-ended question that asked for a description of a pleasant recent experience. Items from the EPPS were included in order to control for the effects of social desirability. Otherwise, the low-similarity other might be less liked because she was a social deviant, rather than because she held attitudes different from the subject.

On his second visit, 1 week later, the experimenter explained that one purpose of the Personal Inventory was to study how people form judgments about another person from knowing the other person's answers to the Inventory. The experimenter told the subjects that each of them would receive a Personal Inventory filled out by a student in another class. The subjects were instructed to read the other person's answers to the Personal Inventory and then to describe her using the description schedule attached to the back of the Personal Inventory. Actually, the subjects were given a Personal Inventory to read which had been previously filled out by the experimenter. Subjects in the high-similarity conditions received a Personal Inventory with answers to each of the 50 items exactly the same as their own. Subjects in the low-similarity conditions received an Inventory with answers that were exactly opposite to their own; on the items the subject answered A, the other person's item was marked B, and vice versa.

² Thanks are due to Gary Isaacson for his assistance in making his classes available for the study.

The information about fortune was introduced in the answer to the question concerning a pleasing recent experience. For subjects in the good-fortune conditions, the answer to the question was the following:

I have just been notified that I've won a drawing in a department store near home, and the prize is a trip to Paris accompanying the store's New York buyer to see all the spring fashion showings as well as attending private parties and receptions given by the designers. It sounds so wonderful. I can hardly wait to go.

For subjects in the average-fortune conditions the answer was:

I can't think of any one particular experience as most pleasing or enjoyable. Nothing unusual or extraordinary has happened to me recently. Just being here at college has been in general my most enjoyable experience.

The interpersonal description schedule contained 21 personal adjectives: annoying, clever, cynical, decisive, defiant, fashionable, happy, lazy, lucky, meddlesome, modest, naive, nervous, open-minded, responsible, shallow, superficial, sympathetic, thoughtful, vain, and worthy. The subject was asked to indicate how applicable each adjective was to the other person by circling a number on a scale from 0 (extremely inappropriate) to 20 (extremely appropriate) located below the adjective. The interpersonal description schedule included three additional items: "How much do you like this person?" "How similar to yourself is this person?" and "How much would you like to be this person?" These items were answered on scales from 0 (not at all) to 20 (very much). The subjects were told that their ratings of the other would be kept in strict confidence.

After the subjects had completed their ratings of the other, they were asked to write a short paragraph about their reactions to the study. None of the reactions indicated suspicion of the instructions. Finally, the experimenter revealed the true purpose of the study and asked the subjects not to discuss it with anyone.

TABLE 1
MEANS OF RATINGS OF THE OTHER

Similarity	Fortune	Ratings for				
		Similarity	Lucky	Fashionable	Happy	Desire to be the other
High	Good	16.19	14.09	16.24	16.86	12.28
High	Average	13.86	10.23	14.42	13.95	12.00
Low	Good	4.95	12.85	12.52	11.90	2.48
Low	Average	5.09	9.24	10.00	12.48	4.14

RESULTS

In order to have an equal number of subjects in each of the four experimental conditions, 2 persons in the high-similarity-average-fortune condition were randomly eliminated, leaving 21 subjects in each condition.

A check on the manipulation of the similarity of the other was provided by the ratings on the question, "How similar to yourself is this person?" Each subject's score was the same as the scale value she circled; higher scores indicating greater perceived similarity. The means for each condition for this measure are presented in Table 1. It can be seen from Table 1 that the means for the high-similarity conditions were considerably higher than the means for the low-similarity conditions. An analysis of variance of these data showed that the main effect of similarity was significant beyond the .001 level. Neither the effect of fortune nor the interaction approached significance. The manipulation of similarity seems to have been quite effective.

The evidence concerning the manipulation of the fortune of the other is less clear-cut. The adjectives lucky, fashionable, and happy, and the question "How much would you like to be this person?" were included in order to assess some of the possible effects of the fortune manipulation on the impression of the other. The means for these measures for the experimental conditions are presented in Table 1. As can be seen from Table 1, the other was considered more lucky in the good-fortune conditions than in the average-fortune conditions. Analysis of variance revealed that this difference was significant beyond the .01 level. Although the other was rated as somewhat more fashionable in the good-fortune conditions, the effect of fortune did not reach significance ($p < .08$). The ratings of how happy the other was and of the desire to be the other were not significantly affected by the fortune manipulation. Ratings for all of these measures except lucky were significantly greater ($p < .001$) in the high-similarity conditions than in the low-similarity conditions.

TABLE 2
MEANS OF ATTRACTION RATINGS

Similarity	Fortune	
	Average	Good ^a
High	13.06	15.42
Low	11.04	10.35

TABLE 3

ANALYSIS OF VARIANCE OF ATTRACTION RATINGS

Source	MS	F
Similarity (A)	263.38	34.07***
Fortune (B)	14.56	1.88
A × B	48.98	6.34**
Error	7.73	

** $p < .02$, $df = 1/80$
*** $p < .001$, $df = 1/80$.

For none of the measures was the interaction between fortune and similarity significant.

Attraction, the dependent variable, was measured by the average of the mean of the ratings for adjectives (happy, fashionable, and lucky were not included in this summation) and the rating on the item, "How much do you like this person?" Scores for the favorable adjectives were the same as the subject's rating; scores for the unfavorable adjectives were obtained by subtracting the subject's rating from 20. The score on the direct measure of liking was the same as the subject's rating. The decision to average the mean of the 18 adjectives with the score on the direct measure of liking was based on the high correlation between the two ($r = .81$, $N = 84$).

The means of the attraction scores for the experimental conditions are presented in Table 2. Examination of Table 2 shows that the means of the attraction scores for the high-similarity conditions were greater than those for the low-similarity conditions for both the average- and good-fortune conditions. The effect of similarity was significant beyond the .001 level as determined by an analysis of variance which is summarized in Table 3. The effect of fortune was not significant. It can also be seen from Table 3 that the interaction between similarity and fortune was significant at the .02 level.

However, the interaction is in the opposite direction from that hypothesized. It was expected that the difference between the high- and low-similarity conditions would be less in the good-fortune conditions than in the average-fortune conditions. In fact, the difference between the high- and low-similarity conditions was greater in the good-fortune conditions than in the average-fortune conditions. Looking at the results from another angle, low similarity-good fortune was equal to low similarity-average fortune, but high similarity-good fortune was significantly greater than high similarity-average fortune ($p < .01$).

DISCUSSION

The results clearly do not support the hypothesis that the tendency to like similar others more than dissimilar others will be smaller when the other has very good fortune than when the other has only average fortune. However, this does not necessarily mean that Heider's analysis of envy is incorrect. Following the comments quoted in the introduction he also said,

At the same time envy is not a necessary consequence of the inequality of the fortunes of people who are close. For one thing, the specific x which o has received may not have the same personal relevance as m as p is concerned. The mother may not be envious of her son's military distinction because she has no aspirations in this direction. Furthermore, the sentiment relation may be sufficiently strong to control the reaction. Sympathetic pleasure in a basic "we-group" rather than envy may be evoked [Heider, 1958, p. 287].

It is doubtful whether any of the subjects had hopes of winning a free trip to the Paris fashion showings. Perhaps for this reason the good fortune of the other did not have much personal relevance for them. If the subjects in the good-fortune conditions did not compare their own fortune with the other's fortune, envy would not be expected to occur. This might account for the lack of support for the hypothesis.

Of course, it is still necessary to explain why the similar other with good fortune was liked more than the similar other with average fortune, while the dissimilar other with good fortune was not liked more than the dissimilar other with average fortune. One possible kind of explanation is related to Heider's remark that sympathetic pleasure rather than envy may be evoked. The good fortune of the similar other may have been a source of pleasure for the subjects. The similar other was liked, and people tend to feel pleasure at the happiness of someone they like. Or, pleasure may have resulted from a tendency to empathize with the emotions of a similar other (Stotland & Dunn, 1963). In this connection, it is appropriate to mention that the other was rated more happy in the similar-good-fortune

condition than in the similar-average-fortune condition (p. 41). This fact, however, must make the additional assumption that the pleasure created by the happiness of the similar other with good fortune produced an increase in liking for her.

A different kind of explanation is suggested by another statement of Heider's that, "One is more likely to envy a person whom one dislikes [1958, p. 289]." This interpretation would assume that envy did occur but that it occurred in the dissimilar-good-fortune condition where the other's dissimilarity had caused her to be relatively disliked. It must also assume that the good-fortune manipulation tended to increase liking for the other, perhaps because of a kind of halo effect. Such a tendency could explain why the similar other with good fortune was liked more than the similar other with average fortune. Dislike produced by envy of the dissimilar other with good fortune could have offset the general tendency to like people with good fortune, so that the dissimilar other with good fortune was not liked more than the dissimilar other with average fortune.

REFERENCES

- BYRNE, D. Interpersonal attraction and attitude similarity. *Journal of Abnormal and Social Psychology*, 1961, 62, 713-715.
- BYRNE, D., & NELSON, D. Attraction as a linear function of proportion of positive reinforcements. *Journal of Personality and Social Psychology*, 1965, 1, 659-663.
- EDWARDS, A. L. *Edwards Personal Preference Schedule*. New York: Psychological Corporation, 1953.
- HEIDER, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.
- JONES, E. E., & DAUGHERTY, B. N. Political orientation and the perceptual effects of an anticipated interaction. *Journal of Abnormal and Social Psychology*, 1959, 59, 340-349.
- SMITH, A. J. Similarity of values and its relation to acceptance and the projection of similarity. *Journal of Psychology*, 1957, 43, 251-260.
- STOTLAND, E., & DUNN, R. E. Empathy, self-esteem, and birth order. *Journal of Abnormal and Social Psychology*, 1963, 66, 532-540.

(Received February 3, 1966)

RECALL OF NONSENSE AND ATTITUDINAL RIGIDITY

BERNARD PYRON

AND

JOHN KATZ, JR.

University of Wisconsin

University of Illinois

Subjects tended to recall more nonsense sentences than in fact. They also attempted to recall the congruence element of each sentence, even though not the stimulus element. The tendency was pronounced in such a way that response elements were brought into working memory. As a result, recall was made to interfere with the learning of some sentences from the incongruent nonsense sentences. Recall of both kinds of nonsense was poorer in subjects low in attitude rigidity than in high attitude rigidity. Subjects high in attitude rigidity tended to recall significantly more congruent nonsense sentences than subjects rated high in rigidity. Subjects rated high in attitude rigidity recalled significantly more congruent nonsense elements than those who scored low in attitude rigidity. The F score was not significantly greater in recall of nonsense sentences.

The hypothesis was tested that subjects who score low in attitudinal rigidity would learn nonsense sentences significantly more than subjects high in rigidity. Nonsense statements were created by joining elements of sentences in incongruous ways. This learning task was similar to a paired-associate list except that the stimulus and response elements (S-R elements) were composed of sentence fragments rather than single words. The combined S-R elements were not only low in initial associative connection but were meaningless in a realistic sense. Some of these nonsense statements were, it is suggested, imaginative images which might have evoked a reaction of humor, shock, wonder, erotic interest, or a less clearly defined emotion.

Some subjects, we suggest, enjoy dealing with nonsense, especially when it is amusing or imaginative. The history of nonsense in literature (Forster, 1962) supports this notion. But not all nonsense created by joining elements of sentences in incongruous ways is amusing or imaginative. Therefore, an attempt was made in this study to differentiate interesting from uninteresting nonsense.

ATTITUDINAL RIGIDITY: REJECTION-ACCEPTANCE OF ATTITUDE POSITIONS AND STIMULI TENDING TO THREATEN THE PERCEPTUAL ORDERING AND BELIEF SYSTEMS

This study was concerned with rigidity as an attitudinal characteristic rather than with rigidity as behavioral perseveration (Chown, 1959). Attitudinal rigidity is defined as the desire to

maintain either the entire perceptual organization system intact and resistant to change or to maintain certain areas of the system in an unchanging condition. The perceptual organization system is made up of perceptual mechanisms for filtering and organizing the multitude of sensory input that the organism receives (Pyron, 1966). These are the mechanisms by which the subjects relate to their environment. Attitudinally rigid subjects tended to reject stimuli which would threaten to change parts of their perceptual organization systems. Thus to protect their perceptual systems, rigid subjects tended to reject a greater range of stimuli than nonrigid subjects.

It was hypothesized that rigid subjects would reject nonsense sentences more than nonrigid subjects, since irrational and novel combinations of familiar elements might threaten the rigid subject's perceptual organization system. Rigid subjects may fear the suspension of the rational and conventional perceptual organization mechanisms which is necessary to perceive an imaginative relationship between two remotely associated elements.

On the other hand, it is suggested that subjects who are nonrigid accept change in their perceptual organization systems and tend to be open to new stimulus combinations which invite imaginative participation. Hence, nonrigid subjects will perceive more clearly than rigid subjects the nonsense images and imaginative relationships of interesting nonsense. Rigid subjects will tend not to differentiate the interesting from the uninteresting nonsense statements and will recall interesting and uninteresting nonsense equally well. Rokeach (1960) has suggested that closed

¹We wish to thank Tom Parker of the White-water State University Art Department for his help in creating the sentences.

or read subjects make fewer differentiations between stimuli within their range of rejection. It is also hypothesized that the predicted difference in recall of nonsense statements between those high and low in attitudinal rigidity will be greater for the interesting than for the uninteresting nonsense.

METHOD

Scale Values of the Nonsense Sentences

Forty sentences were initially developed by the present writers with the help of another individual (see Footnote 1). Ten sentences were devised for each of four types of S-R combinations, common-interesting (interesting and meaningful), remote-interesting (interesting nonsense), common-uninteresting (meaningful but uninteresting), and remote-uninteresting (uninteresting nonsense). The first half of each of the sentences, the stimulus element, ranged in length from four to seven words, while the second, or response element, ranged in length from one to three words. Then the 40 sentences, composed of S-R elements, were given to two separate groups of students from an introductory psychology class. One group ($N=42$) rated the sentences for interest, and the second group ($N=48$) rated them for remoteness (rational meaningfulness). The method of equal-appearing intervals (Thurstone & Chave, 1929) was used to determine the scale and Q values. Since the authors wanted judgments to tend toward a dichotomous distinction between the remote and common

and the interesting and uninteresting, only four sorting categories were used. The higher the scale value the greater the judged commonness or interest. The higher the Q value, the greater was the disagreement between judges.

Sentences with commonness values of 1.82 and below and interest values of 1.86 and above were selected for the 10 remote-interesting sentences. Another group of 10 sentences with commonness values of 1.81 and below and interest values of 1.93 and below were selected for the remote-uninteresting list. Remoteness and interest values for the remote-interesting and remote-uninteresting sentences are reported in Table 1. This study was concerned only with recall of remote-interesting and remote-uninteresting sentences.

Procedure

Three learning trials were given to all subjects, who were presented the task individually. For the first half of Trial 1, each subject listened to a tape recording of all 20 complete sentences, composed of S-R elements. Then, for the second half of Trial 1, the subject heard only the stimulus elements of each of the 20 sentences and attempted to verbalize to the experimenter the exact response element of each sentence. This same procedure was repeated for Trials 2 and 3. Only when the subject gave the exact response element was his response scored as correct. Since all subjects were given three trials,

TABLE 1
COMMONNESS AND INTEREST SCALE VALUES

Associations	Commonness scale value	Q	Interest scale value	Q
Remote-interesting				
1. The November night: disrobes for me	1.82	1.5	3.52	1.0
2. She rubbed brandy into the eyes of: the piano	1.10	.6	2.60	1.8
3. Apollinaire is standing on my: eyelids	1.30	.8	3.31	1.5
4. From his whiskers, John Brown combed: friendly historians	1.25	.8	3.00	1.9
5. Henry kicked out the bottle of: correlation coefficients	1.24	.7	2.20	1.5
6. A good cigar is hanging from the: houses going by	1.26	.8	1.86	.9
7. The eternal mountain: pokes laughing comb	1.14	.7	2.82	1.1
8. I smell some: purple hanging faces	1.25	1.0	2.10	1.4
9. The climax invents itself through: goat-hair brushes	1.11	.5	2.93	1.2
10. A cure for asthma is: a crooked stick	1.22	.8	2.51	1.8
Mean scale values	1.27		2.68	
Remote-uninteresting				
1. The proper use of lined paper reduces: heat loss	1.52	1.1	1.93	1.8
2. The distortion of optical lenses is: unmotivated reaction	1.20	.9	1.52	1.2
3. He read the book rapidly with: immediate color	1.53	1.0	1.80	1.7
4. Go through the door which is of: similar pitch	1.81	1.4	1.49	1.3
5. The back orders were inside the: residual conclusion	1.26	.8	1.22	.6
6. Coefficients reduce the: heating establishment	1.29	.8	1.35	.7
7. Modern hospitals preserve the: geographical strata	1.61	1.2	1.47	1.2
8. Cost analysis was perfected for: intransitive verbs	1.15	.6	1.16	.6
9. The shape of the building was: infrequent	1.28	.7	1.82	1.5
10. Crop rotation is essential to: operating apartments	1.27	.9	1.90	1.1
Mean scale values	1.39		1.56	

scores for each of the two sentence conditions—remote-interesting and remote-uninteresting—could range from 0 to 30.

Subjects

Sixty college freshmen, 30 males and 30 females, from an introductory psychology class at Wisconsin State University, were given the three personality tests and the verbal learning task. This was a separate group from those who judged the commonness and interest of the sentences. All 60 subjects were presented with both remote-interesting and remote-uninteresting sentences. Each of the three different orders of presentation of the sentences, randomly determined, was given to a group of 12 subjects.

Three Individual Difference Measures

The three variables which loaded above .50 on the attitudinal rigidity factor in a previous study (Pyron, 1966) were given to all subjects. These three variables were as follows:

1. *Dogmatism Scale.* The 11-item Form F of Rokeach's (1959) Dogmatism Scale was used and the subjects were provided with a 4-point scale with six verbal anchors for each statement. The higher the score, the greater the dogmatism or closedness of the belief system.

2. *Change Inventory.* The 18 items of the Change Inventory (Pyron, 1965, 1966) fall into the following groups: (a) a desire for order, certainty, and a knowledge of what one will be doing in the future, (b) flexibility, or the ability to change when faced with new situations, (c) acceptance of change in society, (d) sensitivity to change in the natural world, (e) the desire to live for the present rather than the future, (f) affirmation of new experiences which change one's outlook, and (g) becoming, or the value of change toward self-fulfillment. Each statement was rated along an 18-point scale with six verbal anchors. To control for possible yeasaying and naysaying response sets, the subjects were instructed both to agree with and to disagree with nine items. Half of the items were worded in the direction of acceptance of change (positive) and half in the direction of rejection of change. Agreement with a positive statement was scored, and disagreement with a negative statement was also scored. The higher the score, the greater the acceptance of change.

3. *F Scale.* The 28-item version of the original F Scale (Adorno, Frenkel-Brunswick, Levinson, & Sanford, 1950) with six scale positions and six verbal anchors for each statement was used. The higher the score, the greater the authoritarianism.

RESULTS AND DISCUSSION

To test the hypothesis that those subjects who were nonrigid would recall significantly more nonsense sentences than those who were rigid, the group of 60 subjects was divided into four groups on each personality measure, with 15 subjects in each group. Three separate 2 × 4

repeated-measures analyses of variance were performed (Duncan range tests (Duncan, 1955)) since each subject had three measures of the three personality measures. When a measure failed to reach group significance at a .05 level, it was used in subsequent analyses. The correlations between each of the three attitudinal rigidity measures and recall of remote-interesting and remote-uninteresting sentences

The Three Measures

1. *Dogmatism.* Only the Dogmatism Scale was significantly related to recall of the two nonsense conditions as a four-category variable ($F = 3.81, p < .05$). The interaction between dogmatism and interest of sentences of 1.91 fell short of significance at the .05 level. The correlation between dogmatism and recall of remote-interesting associations was $-.36 (p < .01)$, while the correlation between dogmatism and remote-uninteresting sentence recall was $-.16$ (not significant). Table 2 shows the means of the four dogmatism groups for remote-interesting and remote-uninteresting sentences. Subjects who scored low on dogmatism tended to learn the remote-interesting sentences significantly more than those who scored high in dogmatism. The dogmatism boys also tended to score higher in recall of remote-uninteresting sentences, but the difference between the means of the highest and lowest 25% of the subjects was not significant by Duncan range tests (Table 2). The hypothesis that the subjects who scored highest in dogmatism must rigid would recall interesting and uninteresting nonsense equally well was confirmed (Table 2).

2. *Change Inventory.* The Change Inventory was not significantly related to recall of the two nonsense conditions as a four-category variable. But when the subjects were dichotomized on the basis of their scores on the Change Inventory into the highest and lowest 50%, the overall F was 4.91, $p < .05$. Again, the interaction between acceptance of change and interest of sentences was not significant. The Change Inventory correlated .30 ($p < .05$) with recall of remote-interesting sentences, and .19 (not significant) with recall of remote-uninteresting sentences. Subjects who scored high in acceptance of change learned the remote-interesting sentences significantly more than those who scored low on change (Table 2). Subjects who scored high in acceptance of change also learned the remote-uninteresting associations more than those who scored low, but the mean difference between

TABLE 2

MEANS FOR DOGMATISM AND ACCEPTANCE OF CHANGE
FOR REMOTE-INTERESTING (R-I) AND REMOTE-
UNINTERESTING (R-U)

Variable	R-I	R-U
Dogmatism		
Lows (least dogmatic)	12.3	9.1
Low mediums	10.3	8.3
High mediums	9.3	6.7
Highs (most dogmatic)	7.8	7.5
Difference between highs and lows	4.5**	1.6
Acceptance of change		
Lows (most rejecting)	8.8	7.0
Highs (most accepting)	11.3	8.8
Difference between highs and lows	2.5*	1.8

Note.—Overall $F = 3.81$, $p < .05$, for Dogmatism. Overall $F = 4.91$ (as a dichotomous variable), $p < .05$, for Acceptance of Change.

* $p < .05$.

** $p < .01$.

high and low groups was not significant (Table 2). The hypothesis that those who scored lowest in acceptance of change would recall interesting and uninteresting nonsense equally well was not confirmed (Table 2).

Dogmatism and acceptance of change were related to recall of both interesting and uninteresting nonsense. However, there was a stronger relationship between those two attitudinal rigid-

ity variables and recall of interesting nonsense than for recall of uninteresting nonsense.

3. *F Scale*. The *F Scale* did not reach overall significance either as a four-category variable or as a dichotomous variable. The four-category *F* was 1.44. The *F Scale* correlated $-.07$ with recall of remote-interesting sentences and $-.11$ with recall of remote-uninteresting sentences.

REFERENCES

- ADORNO, T., FRENKEL-BRUNSWIK, E., LEVINSON, D., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
- CHOWN, S. M. Rigidity—a flexible concept. *Psychological Bulletin*, 1959, 56, 195-223.
- DUNCAN, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, 11, 1-42.
- FORSTER, L. W. *Poetry of significant nonsense*. Cambridge: Cambridge University Press, 1962.
- PYRON, B. A factor-analytic study of simplicity-complexity of social ordering. *Perceptual and Motor Skills*, 1966, 22, 259-272.
- PYRON, B. Intercorrelations of simplicity-complexity of social ordering, accuracy of perception, acceptance of change, self-reliance and sex attitude. *Perceptual and Motor Skills*, 1965, 20, 1293-1304.
- ROKEACH, M. *The open and closed mind*. New York: Basic Books, 1960.
- THURSTONE, L. L., & CHAVE, E. J. *The measurement of attitude*. Chicago: University of Chicago Press, 1929.

(Received April 15, 1966)

Journal of Personality and Social Psychology
1967, Vol. 5, No. 4, 466-468

SUICIDE, HOMICIDE, AND THE EFFECTS OF SOCIALIZATION

DAVID LESTER

Brandeis University

The aggressive behaviors of nonliterate societies, as manifested in their suicide and homicide rates, were investigated to see if they were related to socialization practices in the societies. No association was found. This was seen as casting doubt on the hypothesis put forward by Henry and Short and by Gold of such an association.

Henry and Short (1954) suggested that different techniques of punishment lead children to express their aggression in different ways, and that these styles of expression persist into later life. In particular, children whose punishment involved love-oriented techniques would tend to inhibit aggression and to manifest more self-directed aggression, whereas children who had experienced more physical punishment would have less inhibition of aggression and tend more to direct their aggression outward.

Gold (1958) assumed that females would have experienced more psychological punishment than

males, officers more than enlistees, whites more than Negroes, and urban inhabitants more than rural inhabitants. Therefore, as a result of this differential experience of punishment, males, enlistees, Negroes, and rural inhabitants should manifest less inward-directed aggression than females, officers, whites, and urban inhabitants, respectively. Gold extended his argument to the aggressive behaviors of suicide and homicide. To support his predictions Gold did not use crude suicide and homicide rates, but argued that a more meaningful measure would be the ratio of the suicide rate to the combined suicide and

homicide rate. His predictions were confirmed for sex, race, urban-rural inhabitation, and army personnel.

Palmer (1965) had three judges rate 40 non-literate societies independently for the incidence of suicide and homicide in each. The three ratings were summed to give a rating of 0 through 21 for the incidence of each of these behaviors in each society. Whiting and Child (1953), using a similar technique, rated 75 societies for the relative importance of love-oriented techniques of punishment in the socialization of the children in the society. Thirty-three societies appear in both of these studies. The aim of this paper is to see whether the results of Gold's study can be generalized to these societies.

If the relationship that Gold found is applicable to these cultural data, it would be expected that societies in which love-oriented techniques of punishment are relatively important will display a higher incidence of suicide relative to homicide.

PROCEDURE

The rates for the incidence of homicide and suicide in each of the 33 societies¹ were taken from Palmer (1965). Five measures were compared with socialization techniques: the rating for the incidence of suicide, the rating for the incidence of homicide, the difference between the suicide rate and the homicide rate, the ratio of the suicide rate to the homicide rate, and the ratio that Gold used of the suicide rate to the combined suicide and homicide rate. The ratings of child socialization techniques were obtained from Child.² The ratings of the relative importance of love-oriented techniques of punishment were compared with the measures of aggression by means of the median test (Siegel, 1956), since the ratings did not meet the requirements for parametric statistics.

RESULTS

The associations between the ratings of the importance of love-oriented techniques of discipline and the measures of suicidal and homicidal behaviors in the 33 societies were not significant in any of the comparisons. The largest chi-square value (corrected for continuity) was .30. (A value of 3.84 would be necessary for significance at the .05 level.) One of the comparisons was in the predicted direction, three were in the opposite direction, and one had equal cell frequencies.

¹ The societies used were: Ainu, Alorese, Andamanese, Ashanti, Azande, Bena, Chagga, Comanche, Copper Eskimo, Hopi, Ifugao, Jivaro, Kvakuitl, Lamba, Lapp, Lepcha, Maori, Marquesan, Navaho, Omaha, Papago, Rwala, Samoan, Sanpoil, Tanala, Tao, Thonga, Tikopia, Tiv, Trobriander, Wogeo, Yungar, and Zuni.

² J. W. Whiting and I. L. Child, personal communication, March 1, 1966. Some of the ratings are given in Whiting and Child (1953, p. 245).

TABLE 1

COMPARISON BETWEEN INCIDENCE OF LOVE-ORIENTED TECHNIQUES OF PUNISHMENT AND RATIO OF INCIDENCE OF SUICIDE TO THE COMBINED INCIDENCE OF SUICIDE AND HOMICIDE IN 33 SOCIETIES

		Ratio of suicide rate to combined suicide and homicide rate	
		Above the median	Below the median
Importance of love-oriented techniques of punishment	Above the median	6	9
	Below the median	9	7

Note.—The chi-square value, corrected for continuity, was .30.

The comparison that produced the largest chi-square was that which compared the incidence of love-oriented techniques of discipline with the ratio of the incidence of suicide to the combined incidence of suicide and homicide (the ratio that Gold suggested). However, this association was in a direction opposite to that predicted, and suicide tended to be less common in societies rated high for the incidence of love-oriented techniques of punishment (see Table 1).

DISCUSSION

The negative finding here is of interest, for many other studies have reported a relationship between the type of discipline administered to the child and the direction of its aggressive behavior (McCord, McCord, & Howard, 1961; Miller & Swanson, 1960). These investigations have been primarily concerned with children and adolescents, and the aggressive behavior has usually been of a mild nature (as expressed in the ending of a story in a story-completion test, in doll-play sessions, or in behavior towards teachers). However, the relationship has been confirmed with delinquent boys (Glueck & Glueck, 1950) and psychiatric outpatients (Greenfield, 1959).

Perhaps the negative results stem from the fact that both suicide and homicide are extreme forms of aggressive behavior. Both have severe consequences and may have other etiological determinants than less severe manifestations of aggression, such as doll play. Whatever the cause of these negative results, they cast a doubt upon the generality and the interpretation of Gold's results. A cursory glance at Gold's groups indicates that differences in status could also predict the observed differences. It appears, therefore, that the causes of the differences in the suicide and homicide rates in these societies may lie elsewhere than in the techniques of socialization.

REFERENCES

- GLUECK, S., & GLUECK, E. *Unraveling juvenile delinquency*. New York: Commonwealth Fund, 1950.
- GOLD, M. Suicide, homicide and the socialization of aggression. *American Journal of Sociology*, 1958, 63, 651-661.
- GREENFIELD, N. S. The relationship between recalled forms of childhood discipline and psychopathology. *Journal of Consulting Psychology*, 1959, 23, 139-142.
- HENRY, A. F., & SHORT, J. F. *Suicide and homicide: Some economic, sociological and psychological aspects of aggression*. Glencoe, Ill.: Free Press, 1954.
- MCCORD, W., MCCORD, J., & HOWARD, A. Familial correlates of aggression in non-delinquent children. *Journal of Abnormal and Social Psychology*, 1961, 62, 79-93.
- MILLER, D. R., & SWANSON, G. E. *Inner conflict and defense*. New York: Holt, 1960.
- PALMER, S. Murder and suicide in 40 non-literate societies. *Journal of Criminal Law, Criminology and Police Science*, 1965, 56, 320-324.
- SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.
- WHITING, J. W., & CHILD, I. L. *Child training and personality*. New Haven: Yale University Press, 1953.

(Received March 24, 1966)

Journal of Personality and Social Psychology
1967, Vol. 5, No. 4, 468-472

AUTHORITARIANISM, AGGRESSION, AND STATUS¹

MILTON E. LIPETZ AND PETER G. OSSORIO

University of Colorado

Subjects scoring high on the California F Scale and those scoring low ($N = 94$) interacted with high- or equal-status target persons with and without instigation to aggression. Only the least overt aggression measure (pressure, but not shock intensity or duration) gave consistent results. The major finding was that low authoritarians were more hostile toward high-status targets relative to equal-status targets both under conditions of no instigation and instigation to aggress. Reference to contrasting authoritarian and equalitarian ideologies appears to provide a basis for understanding the results.

This study is concerned with the aggressive behavior of high and low authoritarians directed toward persons of differential status in situations with varied instigation for that behavior. Theoretical considerations presented in *The Authoritarian Personality* (Adorno, Frenkel-Brunswick, Levinson, & Sanford, 1950) would lead to the inference that persons higher in authoritarianism would more readily aggress against lower status persons than higher status persons, whereas persons lower in authoritarianism would make no such status differentiations. A study by Thibaut and Riecken (1955) suggests that the differential aggressive response to status on the part of high authoritarians can be observed only after frustration is introduced. Other research using test-response indexes of aggression (Roberts & Jessor, 1958; Wright & Harvey, 1965) indicates that persons high in authoritarianism show greater hostility toward frustrators low in status than those high in status. No differential response is found for persons low in authoritarianism.

The present research constituted an attempt to study the aggressive behavioral consequences of authoritarianism in relationship to social

status.² The expression of aggression in face-to-face contact may well differ from its expression on a personality test. The social constraints present could prevent more overt manifestations of aggression so that more covert indicators might be necessary to observe any effects.

Prior formulations of the concept of authoritarianism and past empirical findings led to the following hypotheses:

1. High authoritarians evidence a differential aggressive reaction to high- as against equal-status persons only under conditions of instigation, whereas low authoritarians show no differential aggressive reaction to a status dimension.

2. High authoritarians are more aggressive than are low authoritarians toward equal-status aggressors.

3. Low authoritarians are more aggressive than are high authoritarians toward higher status aggressors.

METHOD

Subjects

Form 40-45 of the California F Scale consisting of 28 items was administered to 94 male college stu-

¹ This study was supported in part by financial aid to the senior author from the Council on Research and Creative Work of the University of Colorado.

² As this article was being written, a study by Epstein (1965) appeared that also was concerned with behavioral measures of aggression as related to authoritarianism and status.

dents. The scoring of the scale has been previously described (e.g., Lipetz, 1960). The resultant distribution was divided at the median. The high-authoritarian group of 47 subjects had a mean F-Scale score of 113.02 and an *SD* of 11.50, with a range of 97-147. The low-authoritarian group of similar number had a mean F-Scale score of 78.28 and an *SD* of 12.52, with a range of 41-97.

Measures

In addition to the California F Scale, the Buss-Durkee (1957) Hostility-Guilt Inventory was administered to all subjects in order to obtain an estimate of aggression independent of the experimental manipulations. Subjects were also given a list with rough guidelines for the definitions of the needs of abasement, achievement, aggression, autonomy, and deference. They were requested to indicate on an 18-point scale their self-assessment of the strength of each of these needs. In this way additional information might be obtained about the self-perceptions of high and low authoritarians.

The behavioral measures of aggression were obtained from a 10-key shock machine.³ The machine was so designed that it would yield measures of shock intensity, duration of shock, and, through a series of strain gauges, the amount of pressure exerted on the keys. It was believed that the most overt expression of aggression involved the selection of shock intensity, with the duration of time of shock a somewhat less obvious manifestation of aggression. The pressure exerted was considered the least overt expression of aggression, but one that would relate to aggression felt without its direct expression.

Status Differentiation

Subjects were randomly assigned to one of four experimenters.⁴ Two of the four were professors who, through manner, dress, age, and titles used, tried to establish a high-status condition. The other two experimenters through similar devices attempted to create a lower status condition.⁵

Procedure

Subjects filled out the various questionnaires and rating forms in one group session. At a later time each subject was seen individually by one of the four experimenters. Subjects were informed that the experiment was concerned with how strong habits may be changed and how outside interference affects such change. Toward that end they were to write the alphabet backwards under several variations while being timed. After these tasks were completed

the subject was told of the interest in the physiological reactions to interference on this same task and was shown a polygraph that would record those reactions. Electric shock administered by a machine constituted the interference. The subject was told that he would be the experimenter, and that the experimenter would act as the subject because of the interest in the effects of shock over an extremely large number of trials. The machine was explained and the subject given sample shocks up to the fourth key so that he knew that the intensity increased from one key to the next. The experimenter attached an electrode to his wrist and stated that he would try writing the alphabet backwards in the ways the subject had done. A partition separated the experimenter from the subject. The subject was told that as soon as the experimenter started to write he (a) was to give a shock each time the bulb in front of him flashed, which would be every 5 seconds, (b) was to push any key he wished, but only one at a time, (c) could hold any key down for up to 5 seconds, but had to release it each time before the bulb flashed, and (d) was to continue to give the shocks until told to stop. Total time was recorded with a stopwatch which the subject was shown.

The subject was allowed to proceed until 30 shocks were administered. Needless to say the power was turned off. A bank of lights informed the experimenter which shock key was being pushed so that appropriate behavior could be simulated behind the partition. A Grass polygraph kept records of which keys were pushed, time duration, and pressure exerted. The experimenter up to this time was in no way abusive or aggressive, and this constituted what was designated as the "no pull" for aggression condition.

Following this interaction the experimenter went to the polygraph and asserted to that same subject that it had not been operating properly. He then "repaired" it and pointed out that there was sufficient time to repeat the procedure. The procedure was identical except that while the subject was writing the alphabet in various ways the experimenter made disparaging remarks about the subject's motivation and performance and his value as a participant. In order to achieve standardization of the instigating stimulus, within realistic limits, the four experimenters were trained on a prepared script and worked from it during the experiment. This unfair treatment on the part of the experimenter constituted the "pull" for the aggression condition.

RESULTS

Interrelationships with Authoritarianism

A low but significant positive correlation ($r = .25, p < .05$) was obtained between the Buss-Durkee Inventory and the California F Scale. Pearson product-moment correlations of the self-ratings of abasement, achievement, and autonomy with that of authoritarianism were insignificant. However, authoritarianism correlated .32 ($p < .01$) with self-rating of aggression and .25 ($p < .05$) with self-rating of deference. These results are consistent with the theoretical con-

³ Thanks are due E. Cam Tibbals for the design of this machine.

⁴ The writers are indebted to Ronald Gerber, O. J. Harvey, and Edward Pomeroy for their assistance as experimenters.

⁵ It is recognized that the two "lower" status experimenters can in no way be considered low-status persons. They are imbued with certain status characteristics simply on the basis of their roles as experimenters despite such remarks as "we're just helping out." At the minimum they must be considered as equal in status to the subjects.

TABLE 1

COMPARISON OF MEAN PRESSURES EXERTED FOR HIGH AND LOW AUTHORITARIANS (FS) IN INTERACTION WITH HIGH- AND EQUAL-STATUS PERSONS

Group 1	Group 2	Condition	M 1	SD 1	M 2	SD 2	t	p
HF-HS	HF-ES	No pull	55.71	29.28	57.63	23.14	45	0.25
		Pull	54.61	26.26	58.53	23.51	45	0.54
LF-HS	LF-ES	No pull	68.40	20.02	47.34	31.17	45 ^a	2.71***
		Pull	72.37	20.47	45.21	29.17	45 ^b	3.68***
HF-ES	LF-ES	No pull	57.63	23.14	47.34	31.17	46	1.30
		Pull	58.53	23.51	45.21	29.17	46 ^a	1.74**
HF-HS	LF-HS	No pull	55.71	29.28	68.40	20.02	44 ^b	1.72**
		Pull	54.61	26.26	72.37	20.47	44 ^a	2.56***
HF-HS	LF-ES	No pull	55.71	29.28	47.34	31.17	45	0.95
		Pull	54.61	26.26	45.21	29.17	45	1.16
HF-ES	LF-HS	No pull	57.63	23.14	68.40	20.02	45 ^b	1.70*
		Pull	58.53	23.51	72.37	20.47	45 ^b	2.15**

Note.—HF = high authoritarian, LF = low authoritarian, HS = high status, ES = equal status.

^a 1-tailed test.^b 2-tailed test.* $p < .10$.** $p < .05$.*** $p < .01$.

cepts of authoritarian submission and aggression.

Authoritarianism, Status, and Behavioral Measures of Aggression

None of the measures which involved shock intensity or shock duration yielded significant differences between high and low Fs in any status interaction. Nor were there significant mean differences in shock intensity or duration between persons above and below the median on the Buss-Durkee Inventory. On the pressure measure, however, persons high on the Buss-Durkee Inventory exerted significantly more pressure on the keys than those low on the scale in both the no-pull situation ($t = 2.51$, $df = 92$, $p < .01$) and the pull situation ($t = 2.44$, $df = 92$, $p < .01$). Within-subject directional changes between no-pull and pull were evaluated by a two-tailed sign test for each of four $F \times$ Status groups. Three of these comparisons revealed changes that were not significant, whereas one, low Fs interacting with high status, showed an increase from the no-pull to pull situation significant at the .05 level of confidence. Further, the pressure measure produced discernible differences for certain authoritarianism-status comparisons. The results of these comparisons may be seen in Table 1.

When status of the interactor is not considered in the analysis, high and low authoritarians do not differ in the mean pressure exerted either in the no-pull situation ($t = .13$) or the pull situation ($t = .29$).

Hypothesis 1 finds no statistically significant

support although the single predicted mean difference is qualitatively demonstrated. Although high authoritarians did not react differentially to high and equal status under conditions in which instigation was not present (as anticipated), neither did they respond in a differential way to high- and equal-status instigators (although qualitatively the difference between the means is in the predicted direction). The most striking refutation of Hypothesis 1 was found in the behavior of the low authoritarians.

Prior consideration of the concept of low authoritarianism resulted in the expectation that persons low in F would not make status differentiations. The results strongly contradict this assertion. Low Fs show greater feelings of aggression toward high-status than toward equal-status persons. This can be observed both before and after a period of frustration.

Hypothesis 2 is supported by the finding that high Fs evidence more aggression toward equal-status aggressors than do low Fs.

Hypothesis 3 is also supported by the data.

No predictions were made for the cross comparisons which appear in the last two groupings in Table 1. It appears that low Fs experience more aggressive feelings toward high-status persons than high Fs do toward equal-status persons.

DISCUSSION

The failure of the shock and time measures to show any significant differences between experimental groups is somewhat surprising, even though it had been anticipated that the more overt expressions of hostility toward the experi-

menter would be subject to some inhibition on both normal social grounds and on a situational basis (Orne, 1962), and might therefore not be shown under the experimental conditions. This failure in differential sensitivity of the more overt indexes of aggressive motivation might raise questions about the interpretation of the pressure measure. For example, differential pressure on the shock keys might be attributed to conflict or anxiety. However, such an interpretation does not appear to provide any basis for expecting any systematic directional differences (i.e., more pressure rather than less, or vice versa) unless the conflict or anxiety were itself a consequence of the aggressive implications of administering shock by pressing the keys. Thus, when an analysis of the results in terms of the ideological interpretation presented below reveals a coherent and well-patterned set of differences in the pressure measure for the high- and low-authoritarian groups, the result is altogether consistent with the interpretation of the pressure measure as an index of aggressive motivation which, because of its covert form, escaped the inhibition which operated on the more overt shock and time measures.

Independent empirical support for considering the pressure measure to be an index of aggressive motivation is provided by the significant relationship of this measure with the Buss-Durkee scores. Because exerting more or less pressure on the shock key had no instrumental or communicative value for the subject relative to the recipient of the shock or to any onlooker, the pressure measure is interpreted not as an aggressive response (see Kaufmann, 1965), but rather as expressive or symptomatic of an inclination or tendency to act aggressively toward the experimenter. Evidence that the experimental procedure was successful in generating anger was provided by informal interviews immediately following the completion of the experiment.

Two of the results are contrary to the expectations derived from the existing literature. The first is that high-F subjects did not make the expected differential response to high-status as compared to equal-status instigators. It should be noted, however, that the numerical inequalities were entirely in accordance with the prior expectation, even though statistical significance was not attained. It might be noted that the present procedure of contrasting high versus equal status would be expected to produce weaker effects than the more usual contrast of high versus low status.

The second unexpected finding, and quantitatively the most dramatic one, is that low-F subjects did react differentially to equal-status and high-status experimenters. Epstein (1965, p.

588) reports a similar finding and suggests that "... some low scores on the F Scale may be indicative of an 'authoritarianism of the left' which is characterized by hostility and rejection of groups that appear economically privileged or powerful. . . ." A less venturesome explanation may be derived from the proposition that the standard ideology for American college students, if not for the population generally, is one of equalitarianism.

If this is the case, and low F-Scale scores reflect an equalitarian ideology rather than merely the absence of authoritarianism, the low-F subject would have an enduring reason to eliminate status and power differentials of all kinds. The expression of hostility toward a person in a higher status position would be a way of denying the validity of the status differential between that person and the low-F subject. The successful expression of hostility (i.e., one that he "gets away with") would be a way of actually eliminating that differential either entirely or in part. Thus, the low-F subject would have an enduring disposition to behave aggressively toward higher status persons, but he would also have a special reason to avoid engaging in unsuccessful expressions of hostility toward higher status persons (since that would magnify or emphasize the status differential), and therefore neither the expression nor the inhibition of the aggressive tendency would be reflected as a simple behavioral regularity. This consideration would apply with greater force if the high-status person in question were overtly hostile or manifestly unfair to the low-F subject. Conversely, the low-F subject would have an enduring reason to avoid the exercise of power and authority over peers and an even stronger reason for avoiding this in dealing with persons of a lower status. Other considerations being equal, he would therefore be less hostile with a peer or a lower status person, and he would be inclined to give a peer's hostility an initial presumption of validity rather than reacting with immediate counter-hostility or defensiveness as he might well do with a higher status aggressor. (Thibaut & Riecken, 1955, report that their low-authoritarian subjects were positively nurturant and supportive with the lower status instigators. Interestingly enough, Thibaut & Riecken included some "equalitarian" questionnaire items in their instrument for classifying subjects as high and low authoritarians.)

If low-F subjects are regarded as equalitarians rather than merely nonauthoritarians, expectations as to their behavior would be changed accordingly. In general, we would then predict that the performance of low-F subjects would be the mirror image of the performance of the high-F subjects, rather than merely lacking the distinc-

tive features of the latter. This was indeed the case: low-F subjects did press the shock keys harder for the high-status experimenters than for the equal-status experimenters in the no-pull condition ($p < .01$), and also in the pull condition ($p < .01$), and reacted with decreased pressure to hostile, derogatory comments by the equal-status experimenters, whereas they reacted with increased pressure to similar treatment by the high-status experimenters ($p < .05$).

The finding of this inverse relationship, and the fact that eight out of eight inequalities corresponding to status and instigation parameters can be accounted for by reference to ideological differences between high-F and low-F groups indicate that this conceptualization of authoritarianism and equalitarianism merits further study.

REFERENCES

- ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
- BUSS, A. H., & DURKEE, A. An inventory for assessing different kinds of hostility. *Journal of Consulting Psychology*, 1957, 21, 343-349.

Journal of Personality and Social Psychology
1967, Vol. 5, No. 4, 472-476

- EPSTEIN, R. Authoritarianism, displaced aggression, and social status of the target. *Journal of Personality and Social Psychology*, 1965, 2, 585-589.
- KAUFMANN, H. Definitions and methodology in the study of aggression. *Psychological Bulletin*, 1965, 64, 351-364.
- LIPETZ, M. E. The effects of information on the assessment of attitudes by authoritarians and non-authoritarians. *Journal of Abnormal and Social Psychology*, 1960, 60, 95-99.
- ORNE, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 1962, 17, 776-783.
- ROBERTS, A. H., & JESSOR, R. Authoritarianism, punitiveness, and perceived social status. *Journal of Abnormal and Social Psychology*, 1958, 56, 311-314.
- THIBAUT, J. W., & RIECKEN, H. W. Authoritarianism, status, and the communication of aggression. *Human Relations*, 1955, 8, 95-120.
- WRIGHT, J. M., & HARVEY, O. J. Attitude change as a function of authoritarianism and punitiveness. *Journal of Personality and Social Psychology*, 1965, 1, 177-181.

(Received March 25, 1966)

EFFECTS OF BALANCE, POSITIVITY, AND AGREEMENT IN TRIADIC SOCIAL RELATIONS¹

AROLDO RODRIGUES²

University of California, Los Angeles

8 abstract triadic relations were presented to 44 Ss. In each triad S, another person, O, and an impersonal entity, X, were involved. Ss were asked to indicate on each of 5 90-mm. scales: (a) how unpleasant the situation was to them, (b) how much they wanted to change their feelings towards O, (c) how much they wanted to change their feelings towards X, (d) how much they wanted to see O's feelings towards X changed, and (e) how much they wanted to change their feelings towards O only in the particular situation under consideration, but not otherwise. The results showed that balance, positivity, and agreement clearly affect the amount of tension reported by Ss. It was further revealed that P's willingness to change the P/X bond and to differentiate his feelings towards O is more clearly affected by balance forces, whereas his willingness to change the P/O bond is more clearly determined by positivity forces, and his willingness to see the O/X bond changed is more distinctly dictated by agreement forces. The need for a theory in which various sources of cognitive bias are taken into account is highlighted.

The experimental findings of various authors have clearly shown that the predictions derived

¹ The research reported in this paper stems in part from Rodrigues' (1966) doctoral dissertation submitted to the Department of Psychology at the University of California, Los Angeles. The author wishes to acknowledge the wise guidance and helpful criticisms of Harold H. Kelley throughout the undertaking of the work from which the present paper was derived.

from Heider's (1946, 1958) balance principle are strongly supported whenever a positive feeling exists between P and O, and that the principle meets some difficulties when the triadic relationship has a negative P/O bond (e.g., Davol, 1959; Horowitz, Lyons, & Perlmutter,

² Now at the Universidade Católica do Rio de Janeiro, Brazil.

1951; Jordan, 1953; Ohashi, 1964; Price, Harburg, & Newcomb, 1966; Rodrigues, 1965, 1966). The consistency of the experimental evidence on this matter led Price et al. (1966) to suggest a distinction among balanced, non-balanced, and imbalanced systems. Unambiguously balanced would be the $+++^3$ and $+- -$ triads, unambiguously imbalanced would be the $++ -$ and $+ - +$ structures, and all triads having a negative P/O bond would be considered as nonbalanced.

The present paper investigates the existence of forces derived from positivity³ of the feeling between P and O, and from agreement between P and O regarding X, as they affect the tension experienced by P in triadic social structures. The author believes that the existence of such forces may help to explain the empirical results mentioned above, which are unaccounted for by the balance principle alone.

The experiment to be reported replicated in part Jordan's (1953) study, measuring tension in two ways: asking the subjects to indicate how pleasant or unpleasant a given situation was (as Jordan had), and also asking them to indicate how much they wanted to change the triads in each of the four possible ways of changing imbalance into balance. According to Heider (1958) there are four such ways: changing the P/O bond, the P/X bond, or the O/X bond, and differentiating P's feelings towards O. The rationale for the proposed way of assessing tension is fully reported elsewhere (Rodrigues, 1966, pp. 48-50).

METHOD

Subjects

The subjects were 22 male and 22 female undergraduate students at the University of California, Los Angeles. None of them had completed more than

³ Throughout this paper triads are represented by three signs, the first of which refers to the P/O bond, the second to the P/X bond, and the third to the O/X bond.

four courses in psychology nor taken any previous course in social psychology or personality. To achieve an equal representation of sex, an extra 5 males were excluded at random from the analysis.

Procedure

The experiment was performed in a classroom during a regular meeting of a general course in social psychology, at the beginning of the third week of classes. The material covered by the instructor up to the time of the experiment had no bearing whatsoever on consistency theories in general or balance in particular. Each subject received a booklet of stapled sheets containing some instructions and eight task pages. The instructions read as follows:

On each of the following pages you will find:

(a) a situation described and (b) five scales following it. In each situation two persons and something called X are involved. The two persons are *you* and another person, O. The X is not specified and may be anything at all toward which you and O have a positive or a negative feeling. On the first of the five scales, anchored by the three points "Best," "Neutral," and "Worst," you are to indicate how pleasant or unpleasant the described situation is to you. The scale goes from pleasantness, through a neutral point, to unpleasantness. The right end of the scale, where it says Worst, indicates the highest degree of unpleasantness. *You may make a mark anywhere along the scale to translate as accurately as you can the feeling of pleasantness that you experience.* On the next four scales, which have as anchor points "None at all," "Neutral," and "Very Much," I would like you to indicate how much you would want to see the changes described at the top of each scale take place. Here again you may place your mark anywhere along the scale. Your task is the same for each of the eight situations depicted on the following pages.

In the first two of these four scales, the subjects were asked to indicate how much they wanted to change their feeling toward O, and toward X, in the third scale they were asked how much they wanted O's feeling toward X to change, and, finally, in the fourth scale, they were asked how much they wanted to change their feeling toward O *only* in

TABLE 1
MEAN RATINGS OF TENSION FOR BALANCED AND IMBALANCED TRIADS AS MEASURED BY UNPLEASANTNESS AND CHANGE SCORES

	Balanced triads				Imbalanced triads			
	+++	+- -	- - +	- + -	- - -	- + +	++ -	+ - +
Unpleasantness	20.18	34.73	59.75	54.21	64.57	66.61	61.43	64.37
Change P/O	21.77	24.23	39.70	43.39	48.68	54.09	35.45	32.75
Change P/X	20.61	35.77	34.34	22.70	38.41	27.64	31.27	51.61
Change O/X	21.73	26.29	37.70	50.52	33.18	42.36	71.55	59.61
Differentiate	20.80	25.48	36.16	36.18	41.43	53.82	51.18	44.30

regard to the particular situation being considered, but not otherwise.

The eight task sheets presented the subjects with each of the eight types of signed triads. The eight liking situations were randomized in order for each subject, according to a table of random permutations.

The five scales were scored by placing a millimeter ruler along each of them, and the score corresponded to the mark on the ruler indicated by the subject's response. The lowest point was equal to 10, the middle point was 55, and the highest, 99 millimeters.

RESULTS

Table 1 shows the mean ratings of tension found for each of the eight structures according to degree of reported unpleasantness and degree of reported willingness to change a given structure. Table 2 depicts the summary of the analysis of variance performed for the unpleasantness data of Table 1. Individual comparisons between each pair of means were carried out using the Neuman-Keuls procedure (Winer, 1962, p. 114). The mean for the + + + structure was found to be significantly different from all others at beyond the .01 level, except from the + - - mean. The latter was found to approach a significant difference from all other means ($p < .10$).⁴

⁴ All p values reported in this paper are two-tailed.

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE
(UNPLEASANTNESS SCORES)

Source	SS	df	MS	F
Between Ss	10074.2	43		
Within Ss	206267.0	308		
Structures	87431.5	7	12490.2	31.6***
Residual	118835.5	301	394.8	

*** $p < .01$.

No significant differences were found between any other pair of means.

Similar analyses of variance were carried out for the data obtained with each of the remaining four scales of tension, namely, willingness to change P/O, P/X, O/X, and to differentiate. The obtained F scores were, respectively, 9.8 ($p < .01$), 10.9 ($p < .01$), 13.5 ($p < .01$), 2.4 ($p < .05$). When a significant F score had been obtained for each of the five analyses of variance, Scheffé's (1953) test for multiple comparisons was used. The comparisons made were (a) the four balanced triads versus the four imbalanced ones, (b) the four triads with a positive P/O relationship versus the four with a negative P/O relationship, and (c) the four triads in which agreement was found between P and O (+ + +, + - -, - - -, and - + +) versus the four in which disagreement was found between them (- - +, - + -, + + -, and

TABLE 3

SUMMARY OF MULTIPLE COMPARISONS MADE USING SCHEFFÉ'S TEST

Comparisons	Σa^2	D	D^2	A
Unpleasantness				
Balance vs. imbalance	8	-3877	15018929	42667.4***
Positive P/O vs. negative P/O	8	-2835	8027225	22807.5***
Agreement vs. disagreement	8	-2379	5429571	15424.9**
Change P/O				
Balance vs. imbalance	8	-1843	3396649	9677.9*
Positive P/O vs. negative P/O	8	-3147	9903609	28163.6***
Agreement vs. disagreement	8	-111	12321	35.0
Change P/X				
Balance vs. imbalance	8	-1562	2439844	6931.3*
Positive P/O vs. negative P/O	8	712	506944	1440.2
Agreement vs. disagreement	8	-770	592900	1684.4
Change C/X				
Balance vs. imbalance	8	-3063	9381969	26653.3**
Positive P/O vs. negative P/O	8	715	511225	1452.3
Agreement vs. disagreement	8	-4453	19828209	56330.1***
Differentiate				
Balance vs. imbalance	8	-3173	10107929	28715.7**
Positive P/O vs. negative P/O	8	-1137	1292769	3615.8
Agreement vs. disagreement	8	-1057	1117249	3174.0

* $p < .11$.

** $p < .05$.

*** $p < .01$.

+ - +). Table 3 shows the results of the comparisons made.⁶

DISCUSSION

The unpleasantness data of Table 1 are remarkably similar to those obtained by Jordan (1953), and reproduce the same pattern of results highlighted at the outset of this paper. The analysis of the difference between pairs of means for the unpleasantness scores showed that balance with a negative P/O and balance with a positive P/O bond do not belong together (thus enhancing Price et al.'s proposed distinction among balanced, nonbalanced, and imbalanced systems).

Table 3 shows the effect of balance as well as of other forces in action. The effects of balance, positivity, and agreement can be seen when tension is assessed by the unpleasantness scores and by willingness to change the structure. The influence of positivity, for instance, is clearly shown by the data on willingness to change P/O. Regardless of whether the triad is balanced or imbalanced, P is willing to change his negative feelings toward O. The index of willingness to change P/X is less sensitive to positivity and, among the balanced triads, to agreement forces. This index and that of willingness to differentiate are more clearly affected by balance forces, inasmuch as tension is consistently lower for balanced relations than for imbalanced ones, when assessed by those indexes. The conservative Scheffé test showed a significant effect of balance when tension is assessed by willingness to differentiate, and this effect approaches significance when tension is measured by willingness to change P/X. This fact alone would not warrant the assertion that these two indexes are more clearly affected by balance forces. However, this fact together with the very minute effect of positivity and of agreement shown on Table 3, when tension is measured by these indexes, strengthens such a conclusion. The index of willingness to change

O/X, on the other hand, is quite susceptible to the influence of agreement, as clearly shown on Table 3. This same pattern of results has been found in another experiment in which O was someone actually known by P (Rodrigues, 1966).

These findings together with those reported by Zajonc and Burnstein (1964) and by Rodrigues (1966) strongly force us to underline Heider's (1958) and Jordan's (1961, 1966) position in regard to the consideration of balance as a principle rather than a theory. The analysis of P's willingness to change the bonds of a given triad showed that balance as well as other forces operate together in triadic social relations. Sometimes all these forces point to the same direction (e.g., the + + + structure), and the social relation is experienced as quite comfortable. Other times they operate in opposition (e.g., the - + - structure), and P's feelings are affected by this conflict of forces. The findings here reported call for the elaboration of a broad theory of interpersonal relations in which balance as well as other sources of cognitive bias, such as positivity and agreement, are taken into account. This is, in fact, the strongest suggestion of the present paper.

REFERENCES

- DAVOL, S. H. An empirical test of structural balance in sociometric triads. *Journal of Abnormal and Social Psychology*, 1959, 59, 393-398.
- HEIDER, F. Attitudes and cognitive organization. *Journal of Psychology*, 1946, 21, 107-112.
- HEIDER, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.
- HOROWITZ, M. W., LYONS, J., & PERLMUTTER, H. V. Induction of forces in discussion groups. *Human Relations*, 1951, 4, 57-76.
- JORDAN, N. Behavioral forces that are a function of attitude and cognitive organization. *Human Relations*, 1953, 6, 273-287.
- JORDAN, N. *Imbalance in balance "theory."* Santa Monica, Calif.: Rand Corporation, 1961.
- JORDAN, N. *On cognitive balance*. (Res. Paper p-178) Arlington, Va.: Institute for Defense Analyses, 1966.
- OHASHI, M. Sociometric choice behavior and interpersonal perception in a triad. *Japanese Psychological Research*, 1964, 6, 72-87.
- PRICE, K. O., HARBURG, E., & NEWCOMB, T. M. Psychological balance in situations of negative interpersonal attitudes. *Journal of Personality and Social Psychology*, 1966, 3, 265-270.
- RODRIGUES, A. On the differential effects of some parameters of balance. *Journal of Psychology*, 1965, 61, 241-250.
- RODRIGUES, A. *The psycho-logic of interpersonal relations*. (Doctoral dissertation, University of California, Los Angeles) Ann Arbor, Mich.: University Microfilms, 1966, No. 66-6819.

⁶ Statistical notes: Σd^2 is the sum of the signed coefficients used in the comparisons. We used 1 and -1 as coefficients for each triad of each group in every comparison. D stands for the difference between the tension shown in the four triads of the first group and the four of the second group being compared. A is the sum of squares for the comparison of interest and is equal to $\frac{D^2}{n \Sigma a^2}$. Treatment totals were used for the comparisons.

Following Scheffé's (1953) suggestion that one consider taking .10 rather than .05 as the critical level of significance when using his conservative test, we reported the two values that approached .10.

SATRAJIT DUTTA, B.A., M.A., Ph.D., is an Assistant Professor of Psychology in the Department of Psychology, University of Michigan, Ann Arbor, Michigan 48106.

WIGGERS, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

of Michigan, Office of Naval Research.

(Received May 2, 1966)

Journal of Personality and Social Psychology
1967, Vol. 5, No. 4, 476-481

RETENTION OF AFFECTIVE MATERIAL.

A FURTHER VERIFICATION OF THE INTENSITY HYPOTHESIS

SATRAJIT DUTTA AND RABINDRA N. KANUNGO¹

Dalhousie University

The generality of the intensity hypothesis in retention of affective material was tested. Subjects were asked to rate the intensity of affects associated with experimentally induced success and failure experiences and were tested on retention of their experiences under task- and ego-orienting instructions. Results indicate that under task orientation unpleasantness due to failure is experienced more intensely than pleasantness due to success. Under ego orientation, unpleasantness due to failure is experienced as less intense than pleasantness due to success. Further, regardless of orienting instructions, the retention for failure and success was found to be a function of the perceived intensities of unpleasant and pleasant affects, respectively.

Several explanations have been offered to account for selective retention of affective experiences (Alper, 1946, 1952; Edwards, 1942; Freud, 1924; Taft, 1954). However, one of the explanations have been found quite adequate in explaining and integrating the diverse results stemming from studies on selective retention. Recently, Kanungo and Dutta (1966) have offered an explanation for selective retention of affective experiences in terms of perceived intensity of affect. They have shown that the perceived intensity of affect of stimulus material undergoes transformation when experienced under varying degrees of emotional involvement, and that retention of the stimulus material may directly depend upon the perceived intensity of its affect. According to them, the intensity hypothesis, as stated above, can account for results of studies (e.g., Kanungo & Das, 1960) designed to support the frame of reference explanation advanced by Edwards (1942). Moreover, they have argued that better prediction about selective retention of affective material is possible in terms of the intensity hypothesis than in terms of either the frame of reference or the functional (Taft, 1954) ap-

proach. As a further test of the generality of the intensity hypothesis, the present study aims to find out if the intensity of affect can adequately explain memory for success and failure experiences.

Feelings of success and failure were experimentally induced in the subjects under conditions of task- and ego-orienting instructions. The orienting instructions were similar to those used in earlier studies (i.e., Rosenzweig, 1943). The purpose of the task- and the ego-orienting instructions was to create respectively low and high stress or emotional involvement for the subjects. Assuming that the affects of pleasantness and unpleasantness are necessary concomitants of experiences of success and failure, respectively, it was of interest to see the differential effects of such instructions on the perceived intensity of affect of success and failure experiences and their retention. It was thought, for instance, that ego-orienting instructions used in the present study may put the subject on the defensive. This may result in experiencing pleasantness associated with success more intensely than unpleasantness associated with failure because the intensity of experience of failure is minimized and that of success is magnified. On the other hand, particular conditions of task orientation used in this study may make the subject less defensive. Under these conditions the subject may feel unpleasantness due to failure more in-

¹ The study was conducted when the authors were at the Indian Institute of Management, Calcutta. The authors wish to thank B. Earhard for his critical reading of the manuscript and helpful comments.

knowing that presentation time is constant is to time by recognition and recognition may influence without being equal for the purposes of this experiment. Any such bias to the following is purely hypothetical.

4. Under each presentation, recognition of a task due to failure will be perceived as more severe than presentation of a task due to success.

5. Failure ego sensitivity, consequences of a task due to failure will be perceived as less severe than presentation of a task due to success.

6. More information than presented tasks will be retained under task presentation and more pleasant than unpleasant tasks will be retained under no stimulation.

7. Mean interest of perceived effect of the tasks retained will be higher than those of the tasks not retained. This will hold true for both pleasant and unpleasant tasks, irrespective of the preceding instructions.

METHOD

Materials

Fifty 4" x 1" rectangular colored, irregular abstract designs were prepared. The main colors—red, black, and yellow—were used in each design. An attempt was made to make the designs look as uniform as possible in terms of color, texture, and abstract visual impression. Each design concealed a puzzle picture and contained a very rare and ambiguous figure of an object, primitive in appearance which the subject had to identify within a specified time period. The figures in the pictures were made ambiguous in order to make the subject depend upon the experimenter's report to know for certain whether he was correct or mistaken in identifying a figure. Even if the subject pointed out the real figure, he could not be absolutely sure of his correctness without asking the experimenter. The following procedure was followed to choose some puzzle of matched difficulty.

All 50 designs were presented as puzzle pictures to 20 judges. While presenting each picture to a judge, the name of the object to be identified in it was mentioned, and 2 minutes were allowed to locate it. Whenever the judge pointed out a figure in the picture other than the which was supposed to represent the named object, he was reported to be wrong and was asked to try again. If the judge failed to identify the correct figure within this 2 minute period, it was pointed out to him by the experimenter. Next he was asked to rate the degree of difficulty of the puzzle picture on a bipolar 9-point scale extending from extremely easy to extremely difficult, with weights ranging from 1 to 9. The degree of difficulty for each of the puzzle pictures was determined by taking its mean polarity score. Twenty pictures belonging to the middle range of difficulty, that is, falling within a range of 4-6, were selected for the experiment. Each of them concealed a common object such as a car, fish, bird,

and so on. These were selected so that the names of the pictures were homophony and rhyme.

Design

Three separate conditions will be used, namely, (1) control, (2) group A, (3) group B. Each subject will be assigned to one of these conditions.

Procedure

Subjects were randomly assigned to the groups of a minimum of each. The group (770) received 140 pictures, conditions and one more group (330) received unconditioned instructions. The task involving pictures was as follows:

We ask your cooperation in conducting and participating in a certain experiment. You should realize that this has been prepared by the group. Each picture contains a puzzle picture which is to be identified by you. Your job consists, none of these pictures have to be recalled and remembered to keep our name. All we want is to check the validity of those pictures for being retained for giving your reaction to them.

The experimental instructions were as follows:

We are going to present to you a set of pictures and request you to identify each one. You should expect to identify each with a fair degree of accuracy. With this set, we want to know your responses, whether within the allowed range, and you will be asked accordingly in the final analysis. The best chance of correct picture response will be when you make a specific effort to be identified by you.

After such instructions, each subject of both the groups was given the following instructions regarding the procedure:

Each picture with the name of the object about will be presented to you for two minutes. Your task would be to locate this particular object in the picture. You will be given only one chance of giving your answer. Hence, be sure that you point out a particular item representing the real object only after careful consideration of all possible alternatives. Thus we want to emphasize because more than one item in the picture may appear similar in some way to the named object, you to be correct you have to choose the one that resembles the object most. Please give your answer only when you are asked after the 2 minutes is over. Do not guess and do not hurry that. Even if you feel confident that you have discovered the object, remain silent till you are asked and go on scrutinizing the picture from all possible angles to be doubly sure. If, after carefully scrutinizing the picture for two minutes you are not sure of identifying the said object, you must make a guess, however wild it may seem to you. After you give your answer, you will be told whether you have correctly identified the object or not.

Following these instructions, each subject was presented with each of the 20 pictures for 2 minutes

for identification of the hidden object. To each picture, the subject responded after a 2-minute period, and then he was only informed whether this solution was correct or wrong. When the subject's solutions were reported to be wrong, the correct solutions were never pointed out to him in order to ensure lack of closure in the subject. After reporting the correctness of a solution, the experimenter then presented to the subject a 9-point bipolar pleasant-unpleasant rating scale with verbally labeled categories. The subject was asked to evaluate on a 9-point scale his experience with the specific task that he had just performed either correctly or incorrectly. Such ratings constituted the degree of affectivity that the subject had experienced with a particular puzzle picture.

The order of presentation of 20 puzzle pictures was randomly varied over 60 subjects. Further, a predetermined pattern of affect was induced in each subject by reporting half of his solutions (10 puzzle pictures) to be correct and the other half (remaining 10 puzzle pictures) to be wrong. In trying to do this, the experimenter took advantage of the abstractness and ambiguity involved in each picture. If, for instance, a particular puzzle meant to be a failure item according to the predetermined pattern, even when the supposedly correct figure was identified by the subject, the experimenter reported it as "wrong." If, however, a puzzle was meant to be a success item according to the predetermined pattern, any figure the subject suggested in the picture was reported to be correct. Thus, it was possible, for any single puzzle picture, to induce a sense of success in half of the subjects and a sense of failure in the other half of the subjects in each group.

After the subjects finished identifying the hidden objects and rating their experiences with each of the 20 pictures, retention tests were administered. Half of the subjects in each group were asked to recall the names of objects concealed in the puzzle pictures. The other half of the subjects in each group were asked to recognize and sort out the puzzle pictures that they had seen before from a pile of 40 pictures. This pile included 20 original puzzle pictures presented to the subjects and another 20 extra puzzle pictures matched as much as possible with the original ones in terms of color, texture, and abstract visual impression. Whether it was recall or recognition, each subject was allowed 5 minutes.

RESULTS

Each subject rated his experience with each of the 10 success puzzles (where success was induced) and 10 failure puzzles (where failure was induced) on a 9-point bipolar pleasant-unpleasant scale. The subject's ratings were transformed into ordinal weights of 0, 1, 2, 3, and 4 for neutral, mild, moderate, strong, and extreme points on the scale, respectively. The weights had a positive value when the subject rated his experience as pleasant and a negative value when the subject rated his experience as

unpleasant. For each subject, two mean-intensity ratings were calculated: one for the 10 success puzzles and another for 10 failure puzzles. It was noticed that success puzzles were always rated as pleasant and failure puzzles were always rated as unpleasant by the subjects. Taking an average of these mean intensities (polarity ratings) in the TO group ($N = 30$), it was found that the absolute level of perceived unpleasantness associated with failure puzzles is ($M = 2.55$, $SD = .576$) more intense than perceived pleasantness associated with success puzzles ($M = 2.117$, $SD = .624$). This difference is significant at the .001 level ($t = 6.37$). Similar averages for the EO group ($N = 30$) revealed just the reverse pattern. The perceived intensity of pleasant affect associated with success puzzles ($M = 2.593$, $SD = .620$) is greater than the perceived intensity of unpleasant affect associated with failure puzzles ($M = 1.76$, $SD = .733$). This difference is also significant at the .001 level ($t = 6.83$). These results support Hypotheses 1 and 2.

A $2 \times 2 \times 2$ analysis of variance (Lindquist, 1953, Type III design) was performed on the retention scores of the subjects. The three classifications were orientation (task versus ego), retention tests (recall versus recognition), and rated affect (pleasant versus unpleasant affect of puzzles). The result of the analysis presented in Table 1 revealed two significant effects. One was the main effect of orientation ($F = 8.775$, $df = 1/56$, $p < .01$) and the other was the interaction between orientation and affect variables ($F = 71.395$, $df = 1/56$, $p < .01$). The reasons for obtaining these significant effects can be seen from the total and the mean number of puzzles retained by the subjects under various conditions. An examination of these data as presented in Table 2 indicates that under task orientation retention was better than under ego orientation. This can be best explained in terms

TABLE 1
ANALYSIS OF VARIANCE OF RECALL SCORES

Source of variation	MS	F
Orientation (O)	19.200	8.755**
Retention tests (RT)	0.824	0.376
O \times RT	0.536	0.244
Residual (b)	2.193	
Affect (A)	1.200	1.070
A \times O	80.034	71.395**
A \times RT	3.343	2.982
A \times O \times RT	1.630	1.454
Residual (w)	1.121	

** $p < .01$, $df = 1/56$.

TABLE 2
RETENTION OF PLEASANT AND UNPLEASANT PUZZLES

	TO group			EO group		
	Recall	Recognition	Combined	Recall	Recognition	Combined
Pleasant puzzles						
Total score	85	82	167	92	100	192
<i>M</i>	5.667 ^a	5.467	5.567	6.133	6.667	6.400
Unpleasant puzzles						
Total score	108	102	210	78	64	142
<i>M</i>	7.200	6.800	7.000	4.867	4.267	4.567
<i>t</i>	3.992***	3.471**	5.268***	3.296**	6.250***	6.738***

Note.—Each of the differences has been tested by using within-subject error variance. $N = 15$ in both recall and recognition in TO and EO groups.

^a $p < .01$.

*** $p < .001$.

of the degree of motivational stress experienced by the TO and EO groups. It is quite possible that the EO group faced a more stressful situation, which in turn adversely affected their incidental learning of the puzzle as compared to the TO group. The significant interaction between orientation and affect stemmed from the fact that more unpleasant than pleasant puzzles were retained by the TO group and more pleasant than unpleasant puzzles were retained by the EO group. Even when the mean retention scores were calculated separately for recall and recognition tests in each of the groups, the differences still remain statistically significant (see Table 2). These results strongly support Hypothesis 3.

It may be pointed out that the pattern of differential retention of pleasant and unpleasant puzzles by the two groups (see Table 2) is quite consistent with the pattern of their respective intensities of affect as perceived by the two groups. In the TO group, the unpleasant puzzles were rated as more intense and were retained better than the pleasant ones. On the other hand, in the EO group, pleasant puzzles were perceived as more intense and were retained better than the unpleasant ones. Such consistency suggests a strong positive relationship between intensity of affect and retention.

The difference between the retention of puzzles rated as pleasant and of those rated as unpleasant was found to be nonsignificant (see Table 1). Furthermore, no difference was observed between the two measures of recall ($F = .376$, $df = 1/56$). This might seem to be somewhat unexpected because recognition is generally found to yield a higher retention score than unaided recall. It should be noted, however, that the recognition and recall procedures followed in this experiment make the situation somewhat different. In the recognition test, sub-

jects were asked to select the 20 original puzzles they had been asked to solve from a pile of 40 puzzle pictures. Visually all these pictures were very similar to each other in terms of the colors used and the abstract information they provided. Compared to this recognition test, in the recall test subjects were asked for a free recall of the names of hidden objects in the pictures. These names were frequently used common words of the subjects' language. Thus, it was probably easier for the subjects to recall the common and simple names than to recognize the puzzles from among the very similar abstract designs.

Number of items retained (both recall and recognition measures combined) by both the groups was 706 out of a maximum possible total of 1,200 (each of 60 subjects was exposed to 20 puzzles). Thus, there were 494 nonretained items. The mean intensity of affect of the retained items calculated from the polarity of ratings on the pleasant-unpleasant scale is 2.548 ($SD = 1.051$). This mean is significantly higher than the mean of 1.836 ($SD = 1.111$) calculated for nonretained items ($t = 11.484$, $p < .001$). Similar differences are observed between the intensities of affect of retained and nonretained items within pleasant (success) puzzles, as well as within unpleasant (failure) puzzles irrespective of whether the subject is in the TO or EO group. Table 3 presents the comparisons between the mean intensities of perceived affect of retained and nonretained items. Such comparisons are shown separately for recall and recognition and for pleasant and unpleasant puzzles in each group. For each comparison the mean intensity of affect of retained items is higher than that of nonretained items. Except in one case, the observed differences are statistically significant. The data in Table 3 clearly support Hypothesis 4.

TABLE 3

COMPARISONS OF INTENSITIES OF AFFECT OF RETAINED AND NONRETAINED ITEMS

	Pleasant puzzles				Unpleasant puzzles				Pleasant & unpleasant combined			
	Total	M	SD	t	Total	M	SD	t	Total	M	SD	t
TO group												
Recall												
Retained	85	2.341	1.091	3.173**	108	-2.741	1.161	2.522*	193	2.560	1.158	4.496***
Nonretained	65	1.754	1.137		42	-2.214	1.081		107	1.935	1.137	
Recognition												
Retained	82	2.573	0.898	6.770***	102	-2.715	0.963	3.702***	184	2.652	0.938	7.573***
Nonretained	68	1.632	0.766		48	-2.083	0.997		116	1.819	0.896	
Combined												
Retained	167	2.455	1.072	6.635***	210	-2.724	1.082	4.296***	377	2.605	1.057	8.307***
Nonretained	133	1.692	0.966		90	-2.144	1.040		223	1.874	1.052	
EO group												
Recall												
Retained	92	2.739	0.792	3.066**	73	-2.329	1.060	4.912***	165	2.558	0.942	6.222***
Nonretained	58	2.276	1.030		77	-1.494	1.014		135	1.830	1.092	
Recognition												
Retained	100	2.700	1.005	1.222	64	-1.953	1.138	3.026**	164	2.408	1.121	4.482***
Nonretained	50	2.480	1.082		86	-1.372	1.172		136	1.794	1.238	
Combined												
Retained	192	2.719	0.909	2.983**	137	-2.153	1.114	5.612***	329	2.483	1.038	7.462***
Nonretained	108	2.370	1.060		163	-1.429	1.103		271	1.804	1.180	
TO & EO groups combined												
Retained	359	2.596	0.965	7.143***	347	-2.499	1.129	8.670***	706	2.548	1.051	11.484***
Nonretained	241	1.996	1.064		253	-1.684	1.133		494	1.836	1.111	

* $p < .02$.** $p < .01$.*** $p < .001$.

DISCUSSION

Retention of success and failure experiences seems to be a function of the perceived intensity of their pleasant and unpleasant affects, respectively. Thus, under specific conditions of task and ego-orienting instructions as used in the present study, the memory for failure and success experiences can be adequately explained in terms of the perceived intensity of the respective unpleasant and pleasant components associated with them.

In an earlier study using an interruption technique, Rosenzweig (1943) found that under task orientation subjects recalled more of interrupted than completed tasks. These results were similar to the Zeigarnik (1938) effect. Under ego orientation, however, subjects recalled more of the completed than interrupted tasks, confirming expectations based on the Freudian repression hypothesis. Rosenzweig explained these results by assuming that under task orientation subjects were influenced by a need-persistent drive or a tendency to persist in a task until its completion. Such persistence in the tendency to completion resulted in better recall of interrupted than completed tasks. On the other hand, under ego orientation, the interrupted tasks were perceived as a threat to the

ego. Hence, an ego-defensive drive was operative, resulting in poor recall of interrupted tasks as compared to completed ones. It may be noticed that Rosenzweig's findings regarding differential recall of incompleted and completed tasks under two degrees of emotional involvement are consistent with the pattern of retention for success and failure items in the present study. Such consistency leads one to consider the possibility of explaining Rosenzweig's results in terms of intensity hypothesis. One can assume that if a subject is at all involved in an experiment where the interruption technique is used, he is bound to experience a feeling of personal success with the completion of a task and personal failure with its interruption. The affective components of pleasantness and unpleasantness associated with the experiences of personal success and failure, respectively, may account for the recall patterns observed in Rosenzweig's study. In fact, in his study Rosenzweig (1943) observes a concomitant relationship between the completion of a task and success, and also between the incompletion of a task and failure.

The present findings and those of an earlier study (Kanungo & Dutta, 1966) support the notion that intensity of an affective experience as perceived by an individual (or a group of

individuals) may determine its retention. An experience that has relatively more intense affect than another seems to create a greater impression on the individual and hence is retained better. These studies strongly suggest that any variable that can influence or bring about changes in the perceived intensity of affect of the individual's experiences is also the variable that indirectly affects their selective retention. It is, therefore, important that research directed toward studying the nature of selective retention of affective experience pay more attention to conditions affecting intensities of perceived affect than to conditions affecting individuals' attitudes, values, or motivational systems.

REFERENCES

- ALPER, T. G. Memory for completed and incomplete tasks as a function of personality: An analysis of group data. *Journal of Abnormal and Social Psychology*, 1946, 41, 403-420.
- ALPER, T. G. The interrupted task method in studies of selective recall: A reevaluation of some recent experiments. *Psychological Review*, 1952, 59, 71-88.
- EDWARDS, A. L. The retention of affective experiences: A criticism and restatement of the problem. *Psychological Review*, 1942, 49, 43-53.
- FREUD, S. Repression. In, *Collected papers*. Vol. 4. London: Hogarth Press, 1924. Pp. 84-97.
- KANUNGO, R. N., & DAS, J. P. Differential learning and forgetting as a function of the social frame of reference. *Journal of Abnormal and Social Psychology*, 1960, 61, 82-86.
- KANUNGO, R., & DUTTA, S. Retention of affective material: Frame of reference or intensity? *Journal of Personality and Social Psychology*, 1966, 4, 27-35.
- LINDQUIST, E. F. Design and analysis of experiments in psychology and education. Boston: Houghton Mifflin, 1953.
- ROSENZWEIG, S. An experimental study of "repression" with special reference to need-persistent and ego-defensive reactions to frustration. *Journal of Experimental Psychology*, 1943, 32, 64-74.
- TAFT, R. Selective recall and memory distortion of favorable and unfavorable material. *Journal of Abnormal and Social Psychology*, 1954, 49, 23-28.
- ZEIGARNIK, B. On finished and unfinished tasks. In W. D. Ellis (Ed.), *A source book of Gestalt psychology*. New York: Harcourt, Brace, 1938. Pp. 300-314.

(Received February 14, 1966)

Journal of Personality and Social Psychology
1967, Vol. 5, No. 4, 481-486

SIMILARITY AND ATTRACTION IN SOCIAL CONTEXTS

MELVIN J. LERNER

University of Kentucky

RONALD C. DILLEHAY

Texas Christian University

AND WILLIAM C. SHERER

University of Kentucky

39 male undergraduates anticipated either cooperation or competition with another student. In comparison with 20 other Ss who anticipated no interaction (control), Ss in the experimental conditions construed the characteristics of the "stimulus S" to fit a consistent pattern: (a) The person who anticipates competition desires a certain amount of dissimilarity with his competitor—the cognition of the other person's attractiveness is apparently irrelevant; (b) a certain amount of attraction to the other person arises from the anticipation of cooperation, and the cognitions of similarity are unaffected. It was concluded that the relation between the cognitions of similarity and attraction in these social contexts is not adequately explained by either a balance or cognitive generalization theory.

An area of continued interest in social psychology concerns the effects of a person becoming aware that another person is similar to him in some respect. One theoretical position on this issue postulates a tendency to generalize similarity, so that the perception of one common aspect leads to the creation or expectation of other common aspects (Burnstein, Stotland, & Zander,

1961; Stotland & Dunn, 1962, 1963; Stotland & Hillmer, 1962; Stotland & Patchen, 1961; Stotland, Zander, & Natsoulas, 1961). A second body of theory and research on this issue has grown out of work by Haider (1958), Newcomb (1963), and others (Broxton, 1963; Burdick & Burnes, 1958; Byrne, 1961; Izard, 1963). This work has indicated that the perception of similarity leads

to liking, and liking leads to the creation or attribution of increased similarity.

The important issue which divides these two theoretical positions (generalization of similarity versus balance) is whether changes in similarity must be associated with concomitant changes in attraction to a stimulus person. This issue became salient in the interpretation of the results of an earlier study (Lerner & Becker, 1965), the results of which were interpretable from either theoretical position. Rather than replicating the previous study in which subjects were led to anticipate situations that could be either cooperative or competitive in nature and then required to select a similar or dissimilar other with whom to interact, the study reported here manipulated the anticipation of competitive or cooperative interaction with a standard stimulus person and assessed the perception of similarity and the attribution of attractive qualities. In this way the subject was free to indicate similarity and attraction, permitting an examination of the degree to which these are interrelated and of the effect of anticipated cooperation and competition on each of them.

METHOD

The research questions require two experimental groups (anticipating cooperative or competitive interaction) and a control group (no anticipated interaction), with subjects in each of these conditions providing information on perceived similarity and attraction to a stimulus person.

Upon reporting to the experiment, the subject was told that the experiment's purpose was "to investigate how different types of people interact in different types of situations." (For the subjects in the control condition, the word "respond" was substituted for "interact.") He was also told that he would be performing another task with a different person who was waiting in another room. He was then asked to fill out a Personal Items (PI) questionnaire and a Ways to Live (WL) questionnaire. After handing the subject the two questionnaires, the experimenter told the subject that he was going to take copies of the questionnaires to the other subject, and the experimenter then left the room for 1 minute. Subjects then participated in one of the following three conditions:

1. *Competitive condition.* After completing the two questionnaires, the subjects in this condition were told that the object of the task which they would engage in with the other subject would be to "trap or maneuver the other subject into making an error," and that the subject who trapped the other the greatest number of times would be the winner and would receive a "substantial amount of money."

2. *Cooperative condition.* Subjects in this condition were told that the object of the task was for the subjects to help each other to collect as many

points as they could. They were also told that the subject and his partner would be given some money for each point which the two of them were able to collect and that there was a "fairly good chance" that each of them would be able to earn a substantial amount of money.

The subjects in these conditions were led to believe that one of the things the experimenter hoped to learn from the experiment was the degree to which the subjects' knowledge of the other subject affected their ability to perform the task. They were told that there were two conditions in the experiment. In one condition the subjects received no knowledge about one another before performing the task. In the other condition, which they were supposedly in, each subject was allowed to hear a brief interview of the other subject before they performed the task.

3. *Control condition.* Nothing was said to the subjects in the control condition about any type of interaction with the other subject. After they had completed the PI and WL scales, they were further told that the experiment attempted to determine "the degree to which different types of people are able to predict the thoughts and actions of other people." They were then told that both they and the other subject would be interviewed by the experimenter so that each could form an impression of the other. Then they would fill out the PI and WL scales as they thought the other subject would fill them out about himself.

After these initial instructions, the subjects in all three conditions went through the same procedure. The experimenter threw a switch which supposedly enabled the subjects to hear what they believed was an interview in an adjacent room of the other subject by the experimenter. In fact all subjects heard the same 4-minute taped interview by the experimenter that covered background, interests, and activities of a male student. Two stimulus persons were used in an attempt to vary the base line of attractiveness of the stimulus object. One person (SS₁) was in fact a graduate student in psychology following a predetermined script; the other interview was conducted under unrehearsed conditions with a male student from the subject pool who consented to the use of the interview for experimental purposes (SS₂). The answers given by the first interviewee were bland and noncommittal, creating a somewhat neutral stimulus person insofar as the content of his answers was concerned. The answers given in the uncontrived interview were sparse and unremarkable.

After the interview had ended, the experimenter reentered the room and asked the subject to fill out the PI and WL scales as he believed the other subjects would fill them out, and to rate the other subject on a series of 15 bipolar adjective scales. An additional three items measuring desired social distance were then answered. Finally, the subject responded to four items concerning the effects of the experimental induction.

Having completed these scales and items, the subject was informed that the task to be performed with the other subject was not to occur. His per-

ception of the experiment was explored, and an explanation of the experiment and answers to his questions were provided.

The measure of similarity was based on the difference between responses to the PI and WL scales given by the subject to describe himself and to describe the other subject. The PI scale was concerned with the subject's perception of his own characteristic traits and feelings and his usual type of response to certain social situations.¹ The items contained in the WL scale were excerpts from the Ways to Live scale developed by Morris and Jones (1955).² In the words of the authors of this scale, "The various ways to live represent, for the most part, conceptions of a desirable life as embodied in the main religions and ethical traditions [p. 523]." Each item in the PI and WL scales was responded to on a 6-point scale ranging from "disagree very much" to "agree very much." The difference between the subject's response to each item and the response to the stimulus subject was summed over each scale (PI and WL). These total scores were converted to standard scores and added to yield the measure of similarity used in this study.

The social distance measure was concerned with how much the subject would like or dislike (a) getting to know the other subject personally, (b) having a Coke with the other subject after the experiment, and (c) having the other subject as a roommate. Each item contained six response categories from "dislike very much" (scored 1) to "like very much" (scored 6). The subject's desired social distance was computed by combining the ratings on these three questions.

The measure of attraction consisted of 15 pairs of adjectives, each judged by the subject on a 9-point bipolar scale.³ The sum of the scores on the 15 scales provided the measure of attraction.

The four items that served as a check on the experimental induction were: (a) "What do you imagine the interaction will be like between you and the other subject?" (very competitive = 1, very cooperative = 6); (b) "How do you think most people would feel in the kind of situation you will be in?" (very tense = 1, very relaxed = 6); (c) "How comfortable do you imagine most people would be participating in this kind of experiment?" (very uncomfortable = 1, very comfortable = 6); (d) "How well do you think you will like participating in the experimental task?" (dislike very much = 1, like

very much = 6). Of these four scales, only *c* was completed by the subjects in the control condition.

Sixty-one unpaid male students enrolled in introductory psychology classes participated in the study to satisfy a course requirement. Two subjects were excluded from the sample due to failure to accept the experimental ruse.

RESULTS

The data from the validating questions indicate that the subjects were responding appropriately to the experimental instructions: subjects in the competitive condition anticipated competitive interaction rather than cooperative and were more anxious and tense about what the anticipated interaction might bring them. Nevertheless, all subjects were relatively "pleased" with being in the experiment.

Although the findings concerning SS_1 and SS_2 are presented in the same tables, the data obtained using each of these stimulus subjects were actually collected in sequence and are best understood in that context.

An examination of Tables 1 and 2 reveals that when the subjects anticipated interacting cooperatively with SS_1 there was no appreciable increase in perceived similarity and no increase in attractiveness. There was, however (see Table 3), a significant increase in their desire to reduce the social distance between themselves and SS_1 ($\bar{X}_{coop.} = 16.1$, $\bar{X}_{control} = 13.8$; $t = 2.72$, $p < .02$). On the other hand, the anticipation of competitive interaction led to a decrease in similarity ($\bar{X}_{comp.} = 110.5$, $\bar{X}_{control} = 96.5$; $t =$

TABLE 1
MEAN ATTRACTION TO STIMULUS PERSON AS A
FUNCTION OF KIND OF ANTICIPATED
INTERACTION^a

Stimulus object	Anticipated interaction		
	Cooperation	Competition	No interaction
SS_1	97.7 (9)	97.2 (10)	94.1 (10)
SS_2	95.1 (10)	81.4 (10)	75.0 (10)

Analysis of variance

Source	df	MS	F
Conditions (A)	2	1141.0	4.88**
SS (B)	1	1360.0	5.81**
A \times B	2	411.5	1.76
Error	53	233.8	

^a Higher scores denote greater attraction. The number per cell is given in parentheses.

** $p < .02$.

¹ Two of the 10 items read as follows: "When I get into an argument with someone, I sometimes drop the subject rather than make them angry"; "I sometimes feel that people underestimate my capabilities."

² Two of the 13 items read as follows: "One must avoid dependence upon persons or things; the center of life should be found within oneself"; "Man's future depends primarily upon what he does, not on what he feels or on his speculations."

³ Some of the most highly value-laden dimensions were: (a) likable-unlikable, (b) mature-immature, and (c) irresponsible-responsible.

TABLE 2

MEAN SIMILARITY TO STIMULUS PERSON AS A FUNCTION OF KIND OF ANTICIPATED INTERACTION^a

Stimulus object	Anticipated interaction		
	Cooperation	Competition	No interaction
SS ₁	92.1 (9)	110.5 (10)	96.5 (10)
SS ₂	105.5 (10)	105.4 (10)	105.9 (10)

Analysis of variance

Source	df	MS	F
Conditions (A)	2	233.5	1.22
SS (B)	1	832.0	4.35*
A × B	2	477.5	2.50
Error	53	191.0	

^a Lower scores denote greater similarity. The number per cell is given in parentheses.
* $p < .05$.

2.25, $p < .05$), but had no commensurate effect on either attractiveness or social distance.

As had been expected, SS₂ was perceived overall as less attractive than SS₁ ($F = 5.81$, $p < .02$). The subjects also preferred greater social distance ($F = 18.08$, $p < .001$) and perceived him as less similar to themselves than they did SS₁ ($F = 4.35$, $p < .05$).

The anticipation of cooperative interaction with SS₂ led to a significant increase in his attractiveness ($\bar{X}_{\text{coop.}} = 95.1$, $\bar{X}_{\text{control}} = 75.0$; $t = 2.94$, $p < .005$). Under this condition there was no increase in the perceived similarity of SS₂ or in the desire to reduce social distance.

TABLE 3

MEAN SOCIAL DISTANCE DESIRED AS A FUNCTION OF KIND OF ANTICIPATED INTERACTION^a

Stimulus object	Anticipated interaction		
	Cooperation	Competition	No interaction
SS ₁	16.1 (9)	15.1 (10)	13.8 (10)
SS ₂	13.9 (10)	10.6 (10)	13.0 (10)

Analysis of variance

Source	df	MS	F
Conditions (A)	2	22.85	4.61**
SS (B)	1	89.53	18.08***
A × B	2	19.04	3.84*
Error	53	4.95	

^a Lower scores indicate desire for greater social distance. The number per cell is given in parentheses.
* $p < .05$.
** $p < .02$.
*** $p < .001$.

The subjects in the competitive condition exhibited no decrease in perceived similarity to SS₂ and no significant change in perceived attractiveness. They did, however, show a tendency to increase the desired social distance between themselves and him ($\bar{X}_{\text{comp.}} = 10.6$, $\bar{X}_{\text{control}} = 13.0$; $t = 2.50$, $p < .02$). This effect was created mainly by their reactions to Question 3 concerning his desirability as a roommate.

DISCUSSION

These data overall have negative implications for both conventional balance theory and the theory of cognitive generalization of similarities. The relative mean values of the data for control groups are commensurate with balance theory: the better-liked stimulus person is also perceived as more similar and one with whom P desires less social distance. However, changes in attraction that occur when P is to cooperate with a relatively unattractive O are not accompanied by a concomitant change in perceived similarity. On the other hand, when P is to compete with a relatively similar O, changes in similarity are not accompanied by paralleled changes in attraction. In addition, the anticipation of cooperation with an attractive and similar O leads to an expression of desire for less social distance, but this does not occur with a relatively unattractive and dissimilar O, even though he becomes significantly more attractive. Also, P increases desired social distance in the case of competition with a relatively unattractive and dissimilar O, but not with a similar and attractive O, even though he is perceived as significantly less similar facing competition.

Taken together, however, these results create a rather consistent and somewhat surprising pattern. One clear finding is that the anticipation of cooperative interaction led to no increase in the perceived similarity of the other person. At the same time, there was a striking equivalence in the attractiveness attributed to both stimulus subjects in the cooperative condition (SS₁ $\bar{X} = 97.7$, SS₂ $\bar{X} = 95.1$) even though SS₂ was clearly perceived as relatively unattractive in the control condition (SS₂ control $\bar{X} = 75.0$). The anticipation of competitive interaction seemed to elicit an equally consistent picture, but in relation to similarity. The subjects in that condition perceived both SS₁ and SS₂ virtually equal in dissimilarity (\bar{X} SS₁ = 110.5, \bar{X} SS₂ = 105.5). This occurred in spite of the tendency of subjects to perceive SS₁ as relatively similar to themselves in the control condition.

It is clear that although the findings with both stimulus subjects appear to fit together, they do

not support any available theoretical positions. One possible generalization from these findings is that the anticipation of cooperative interaction leads to an observable increase in the perceived attractiveness of the other person but only if he is "initially" unattractive—not if he is attractive to begin with. What seems to be operating is a desire on the person's part to interact with someone of a particular amount of attractiveness. Some data from Jones and Daugherty (1959) support this notion. They found, but had not expected, that the anticipation of cooperative interaction led to an increase in the attractiveness of a relatively unattractive other (from $\bar{X} = 111.44$ in the control condition to 127.45 in the cooperative), and a decrease in the attractiveness for the attractive other (from 136.74 in the control to 125.91 in the cooperative). Apparently their study revealed the same phenomenon: the anticipation of cooperative interaction led to virtually identical degrees of attractiveness for two initially different others.

These findings are sufficiently intriguing and general to elicit conjectures about what is happening in such situations. It is not surprising that the anticipation of cooperative interaction leads to perceiving one's partner as attractive. A number of theoretical orientations would lead to this prediction. To the extent that P anticipates doing something for O and that O will do something for him, P will find O attractive. Another possibility, based on a more functional approach, is that P comes to like O in this context because liking will facilitate cooperative interaction which will lead to their mutual gain. Why these processes do not lead to increased attractiveness for all others may be due to some factor such as the person's anxiety about the possibility that he may not contribute sufficiently to O's welfare. P would then need to keep his attraction to his partner in check.

What is even more puzzling is why the person seems to require a certain amount of dissimilarity with the other when he anticipates competitive interaction. One possibility is that certain norms in our society concerning competitive interaction might elicit a posture of "friendly enemies" or mutual respect. This might account for the required dissimilarity and the irrelevance of the cognitions of attractiveness—the other person is different, but not necessarily undesirable or unattractive.

To test this idea, further data were collected. Ten subjects were led to believe that they were going to be required to "badger, insult, and emotionally upset" SS₁, to take unfair advantage of him, and that SS₁ could not retaliate or even

know his antagonist. The results of this experimental condition (exploitative) showed that the subjects responded to SS₁ in essentially the same fashion as had the subjects in the competitive condition: SS₁ was perceived as equally dissimilar ($\bar{X} = 107.0$), and attraction was unaffected ($\bar{X} = 92.4$). Obviously, this exploitative condition was not defined in terms of friendly or respected combat, and on the basis of the data it would seem unlikely that the changes in similarity and attraction are attributable to the operation of any set of social norms concerning gentlemanly competition. Again, we are left with no clear explanation as to why the anticipation of competitive or exploitative interaction requires dissimilarity, while leaving attraction largely unaffected. One possible explanation is that it would be extremely dysfunctional for a person to dislike everyone with whom he anticipates competitive or exploitative interaction. It may be true, as Fromm (1955) suggested, that much of the interaction in contemporary society is based on competition and exploitation. That being the case, most of us, to keep our cognitions in balance, would have to dislike our colleagues, siblings, tradespeople, neighbors, etc. The total effect would be a rather miserable life for the person whose cognitions were in balance. Instead, what may be happening is that the person "disidentifies" with, feels somewhat (but not entirely) different from, most others with whom he anticipates a competitive or exploitative relationship. If, in fact, he does engage in exploitative or punitive interaction, he may then allow himself, or be forced, to dislike the victim for the sake of reducing dissonance or achieving balance (Davis & Jones, 1960; Glass, 1964).

REFERENCES

- BROXTON, J. A. A test of interpersonal attraction predictions from balance theory. *Journal of Abnormal and Social Psychology*, 1963, **64**, 394-397.
- BURDICK, H. A., & BURNES, A. J. A test of "strain toward symmetry" theories. *Journal of Abnormal and Social Psychology*, 1958, **57**, 367-370.
- BURNSTEIN, E., STOTLAND, E., & ZANDER, A. Similarity to a model and self-evaluation. *Journal of Abnormal and Social Psychology*, 1961, **62**, 257-264.
- BYRNE, D. Interpersonal attraction and attitude similarity. *Journal of Abnormal and Social Psychology*, 1961, **62**, 713-715.
- DAVIS, K. E., & JONES, E. E. Changes in interpersonal perception as a means of reducing cognitive dissonance. *Journal of Abnormal and Social Psychology*, 1960, **61**, 402-410.
- FROMM, E. *The sane society*. New York: Rinehart, 1955.

- GLASS, D. C. Changes in liking as a means of reducing cognitive discrepancies between self-esteem and aggression. *Journal of Personality*, 1964, 32, 531-539.
- HEIDER, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.
- IZARD, C. E. Personality similarity and friendship: A follow-up study. *Journal of Abnormal and Social Psychology*, 1963, 66, 598-600.
- JONES, E. E., & DAUGHERTY, B. N. Political orientation and the perceptual effects of an anticipated interaction. *Journal of Abnormal and Social Psychology*, 1959, 59, 340-349.
- LERNER, M. J., & BECKER, S. Interpersonal choice as a function of ascribed similarity and definition of the situation. In I. D. Steiner & M. Fishbein (Eds.), *Current studies in social psychology*. New York: Holt, Rinehart & Winston, 1965. Pp. 228-236.
- MORRIS, G., & JONES, L. V. Value scales and dimensions. *Journal of Abnormal and Social Psychology*, 1955, 51, 523-535.
- NEWCOMB, T. M. Stabilities underlying changes in interpersonal attraction. *Journal of Abnormal and Social Psychology*, 1963, 66, 376-386.
- STOTLAND, E., & DUNN, R. E. Identification, "oppositeness," authoritarianism, self-esteem, and birth order. *Psychological Monographs*, 1962, 76(9, Whole No. 528).
- STOTLAND, E., & DUNN, R. E. Empathy, self-esteem, and birth order. *Journal of Abnormal and Social Psychology*, 1963, 66, 532-540.
- STOTLAND, E., & HILLKER, M. J., JR. Identification, authoritarian defensiveness, and self-esteem. *Journal of Abnormal and Social Psychology*, 1962, 64, 334-342.
- STOTLAND, E., & PATCHEN, M. Identification and changes in prejudice and in authoritarianism. *Journal of Abnormal and Social Psychology*, 1961, 62, 265-274.
- STOTLAND, E., ZANDER, A., & NATSOULAS, T. Generalization of interpersonal similarity. *Journal of Abnormal and Social Psychology*, 1961, 62, 250-256.

(Received March 31, 1966)

Journal of Personality and Social Psychology
1967, Vol. 5, No. 4, 486-490

COGNITIVE CONTROL AND COGNITIVE DISSONANCE¹

DAVID L. WOLITZKY

Research Center for Mental Health, New York University

Cigarette smokers and nonsmokers were divided into constricted (high-interference prone) and flexible (low-interference prone) groups based on their performance on the Stroop Color-Word Test. Responses on a smoking questionnaire administered to all Ss indicated that flexible compared with constricted smokers denied a relationship between cigarette smoking and lung cancer to a significantly greater extent, despite the fact that they smoked significantly less. The belief difference was found in the entire sample of smokers as well as in a subgroup of stable smokers. Constricted Ss showed a greater tendency, almost significant, toward recent decreases in smoking. The results suggested that facility in ignoring the intrusive words on the Stroop Color-Word Test extends to the area of belief.

Festinger (1957) defined cognitive dissonance as psychologically uncomfortable tension arising from contradictory cognitions. People are motivated to avoid dissonance and to reduce it when it occurs. A similar emphasis on the tendency towards consonance among cognitions is contained in the concepts of congruity (Osgood & Tannenbaum, 1955) and balance (Heider, 1958). The numerous studies based on dissonance theory (see Glass, in press, for a recent review) have tended to neglect individual differences in

¹ This investigation was supported by a United States Public Health Service Career Development Award (Grant 1-K3-MH-17, 450 and Corollary Grant MH-11073) from the National Institute of Mental Health. The author would like to thank George S. Klein for his helpful suggestions and Leslie Bowling for his assistance in the study.

A version of this paper was presented at the Eastern Psychological Association meeting, New York, April 1966.

the perception and tolerance of dissonance and in the modes of dissonance reduction. A similar neglect was evident in the early days of the "new look" in perception until it was realized that a more precise evaluation of experimental results required attention to individual differences (e.g., Klein, Schlesinger, & Meister, 1951). The present study attempts to link the theory of cognitive dissonance with the concept of cognitive controls. Cognitive controls refer to an individual's relatively enduring modes or strategies of processing information and reacting to situations. They are considered to be developmentally stabilized, conflict-free structures which operate consistently despite shifts in situational conditions. With this conception of cognitive control, Klein (1954) demonstrated that the direction and degree of influence exerted by a need (thirst) upon a variety of perceptual cognitive tasks differed in "constricted" and "flexible" sub-

jects. The constricted-flexible control dimension refers to individual differences in proneness to interference on a color-word task which requires the subject to ignore compelling, irrelevant stimuli.

Interference proneness seems relevant to dissonance studies in which the subject has the option either of ignoring or denying a dissonance-arousing cognition. For instance, high-interference-prone (constricted) individuals should find it hard to avoid or deny an insistent but dissonant cognition in their efforts to maintain cognitive consistency; those less subject to interference (flexible subjects) should be more facile in this respect.

To test this notion, an early study cited by Festinger (1957) was selected as a point of departure. Festinger reported that cigarette smokers, compared with nonsmokers, did not believe that cigarette smoking was related to lung cancer. Such a belief, according to Festinger, would conflict or be dissonant with the smokers' desire to continue smoking and to remain healthy. Some smokers did, however, admit a smoking-cancer link. In more recent studies, the percentage of smokers who reported belief in a smoking-cancer link ranged from 48% (Horn, 1963) to 67% (Baer, 1966). How can one account for these variations? Perhaps smokers are able to deny a smoking-cancer link only to the extent that they are generally able to ignore intrusive, inconsistent information.

It should be noted that the climate of opinion surrounding smoking and cancer has changed considerably since the time of Festinger's study. Mounting evidence and recent legislative action have made the smoking-cancer relationship less ambiguous and more difficult to deny. Therefore, smoking should provoke greater dissonance than it did in the past. Despite these changes, the following hypothesis can still be tested: Smokers who are characterized by constricted cognitive control will admit that a link exists between cigarette smoking and lung cancer, while smokers characterized by flexible cognitive control will tend to deny it.

METHOD

Subjects

Seventy-eight male undergraduates (58 cigarette smokers and 20 nonsmokers) at New York University served as volunteers in individual sessions.

Procedure

Stroop Color-Word Test (CWT). Part I: The testing began with a warm-up page of color names (red, blue, green, yellow) typed in black ink; subjects called out the names of colors as quickly and as accurately as possible. Part II: After a 1-minute rest, the subjects were shown a page of colored strips of asterisks and asked to read aloud the colors as quickly and as accurately as possible.

Part III: The third and main part of the task consisted of color names printed in incongruous colors (e.g., the word "red" printed in blue, "green" printed in yellow, etc.). Subjects were told to ignore the words and to read aloud the colors as quickly and as accurately as possible.

In each of the three sections of the task, the page contained 100 stimuli, 10 to a line. Time scores were recorded as the index of performance (Gardner, Holzman, Klein, Linton, & Spence, 1959).

Smoking questionnaire. Following the CWT, subjects were asked about their smoking habits: (a) whether they smoked cigarettes, cigars, or a pipe,² (b) how many packs a day they smoked currently, (c) whether and by how much they had cut down or increased their smoking within the past year, and (d) the extent to which they believed that a relationship exists between cigarette smoking and lung cancer.³ Degree of belief was expressed on a 4-point rating scale: 1—not at all convinced, 2—partially convinced, 3—pretty sure, 4—certain.

RESULTS

The correlation, based on all 78 subjects, between the time scores on Part II (colors alone) and Part III (color-word incongruence) was .56 ($p < .01$).⁴ Part III scores were therefore adjusted by means of a regression formula in order to obtain an interference score which was independent of color-coding time. The resulting scores were ranked and divided at the median to form constricted (high-interference) and flexible (low-interference) groups. The difference between the two groups with respect to degree of belief in a smoking-cancer link was then assessed. The mean for the constricted group was 3.24 compared with 2.62 for the flexible group ($t = 3.40$, $p < .002$), thereby supporting the hypothesis.

Festinger (1957)⁴ reasoned that heavy smokers had more dissonance to reduce than light smokers. In accord with this assumption, he found that amount of cigarette smoking was directly related to the extent of denial of a smoking-cancer link. It was necessary, therefore, to ascertain that the observed difference in belief was not simply due to different amounts of smoking in the flexible and constricted groups. The mean number of packs smoked each day (questionnaire item b) in the constricted group was 1.16, versus .86 in the flexible group ($t = 3.26$, $p < .002$). The greater denial of a smoking-cancer link found to be typical of

² Cigar and pipe smokers were excluded from the study.

³ Pervin and Yanko (1965) investigated alternative modes of dissonance reduction in smokers. The item that most clearly differentiated smokers from nonsmokers was similar to the above question regarding belief.

⁴ Two-tailed tests of significance were used throughout.

flexible subjects cannot then be attributed to amount of smoking since, in fact, they were the lighter smokers of the total sample.

In addition to dissonance created by amount of current smoking, recent *changes* in the smoking habit might affect the magnitude of dissonance. It has been found, for example, that the more recent the choice between attractive alternatives, the greater the dissonance (Ehrlich, Guttman, Schonbach, & Mills, 1957). Therefore, a recent increase in smoking might raise the level of dissonance, while a recent decrease might lower it. If those who increased smoking were mainly flexible subjects and if those who decreased were mainly constricted subjects, the difference in belief could be attributed to a difference in magnitude of dissonance. In actual fact, when the proportion of decrease in each group is compared, the results show that 41% of the constricted subjects decreased their smoking, compared with 24% of the flexible subjects ($z = 1.72$, $p = .08$, two-tailed test).⁵ Thus, the stronger belief in a smoking-cancer link in the constricted group might be due to the use of an alternative mode of dissonance reduction, namely, decreased smoking.

In order to rule out this possible artifact, subjects whose smoking habits had changed were eliminated from the data analysis, leaving only subjects who reported that their smoking habits had remained *stable* for at least 1 year prior to the study. The respective mean belief score of 3.33 for the constricted group and 2.40 for the flexible group was significantly different ($t = 3.21$, $df = 25$, $p < .001$). As in the original analysis, flexible smokers denied a smoking-cancer link. In this group of stable smokers, constricted subjects still smoked significantly more than flexible subjects (1.29 versus .90, $t = 3.00$, $df = 25$, $p < .05$), again suggesting that while constricted subjects presumably had more dissonance to reduce, they nonetheless expressed a stronger belief in a smoking-cancer link. On the basis of these analyses, the greater denial in flexible subjects cannot be attributed to greater dissonance in these subjects.

Another possible artifact in these results is the possibility of response bias. Perhaps constricted subjects tend to select more extreme points on a rating scale. A check on this possibility was offered by the group of 20 *nonsmokers* who were run through the same procedure. The mean belief rating of the constricted subjects was 2.80, compared with 3.10 for flexible subjects. This is a reversal of the direction found among constricted and flexible smokers and indicates the absence of a response bias.

After having ruled out various possible arti-

⁵ There was no significant difference in the proportion of subjects who increased in each group.

TABLE 1
SMOKING, INTERFERENCE SCORE, AND BELIEF
IN SMOKING-CANCER LINK

	Constricted (high-inter- ference scorers)	Flexible (low-inter- ference scorers)	M
Smokers	3.24 ^a	2.62 ^a	2.93
Nonsmokers	2.80 ^b	3.10 ^b	2.95
M	3.13	2.74	

^a $N = 29$.

^b $N = 10$.

facts, the belief scores of all subjects were subjected to a 2×2 analysis of variance for unequal cell frequencies (Winer, 1962). There was no main effect of constricted-flexible control or smokers-nonsmokers. However, the interaction was significant ($F = 5.22$, $df = 1, 74$, $p < .05$) and again indicates that the constricted-flexible distinction is relevant only among smokers (see Table 1).

The true relationship between the key variables in this study (constricted-flexible control and belief in a smoking-cancer link) might be confounded by the fact that constricted and flexible smokers differ slightly, but significantly, in amount of current smoking and almost significantly in recent decreases in amount of smoking. Partial correlations were computed, therefore, based on the total sample of 58 smokers and on the subsample of 27 smokers whose smoking habits had remained stable for at least 1 year. The original correlations between constricted-flexible control and belief in a smoking-cancer link were .414 ($N = 58$, $p < .01$) and .541 ($N = 27$, $p < .01$). After partialing out amount of smoking as a variable, the comparable correlations dropped negligibly to .381 and .516, still significant at $p < .01$.

It is noteworthy that the correlation between amount of smoking in the total sample of smokers and degree of belief in a smoking-cancer link, is .177, a trend contrary to the straightforward dissonance-theory expectation that heavier smokers would tend to deny a smoking-cancer link. A similar lack of support for the dissonance-theory prediction is seen in the failure of smokers to show greater denial than nonsmokers. The mean belief score of the nonsmokers was 2.95, compared with 2.93 for the smokers (see Table 1).

DISCUSSION

The results support the initial hypothesis: Smokers with constricted cognitive control tend to admit a smoking-cancer link; smokers with flexible control, even though they smoke less, tend to deny such a link. These differences were statistically significant whether one included or eliminated subjects whose smoking habits had

changed within the year prior to testing. In addition, constricted compared with flexible subjects showed a near-significant trend toward decreased smoking in the year prior to study. The pattern that emerges, then, is that flexible smokers smoke less than constricted smokers, deny a smoking-cancer link, and maintain their level of smoking, while constricted smokers smoke more, admit a smoking-cancer link, and tend to cut down on their smoking.

Constricted-Flexible Cognitive Control and Belief in a Smoking-Cancer Link

The results regarding belief differences suggest that belief in a smoking-cancer link is quite closely tied in with a person's ability to deal with discrepant information or intrusive cognitions. If a smoker is asked to report his belief in a smoking-cancer link, a strong belief in such a link constitutes an intrusive idea with respect to his desire to continue smoking and to remain healthy. It is thus an unwanted but compelling cognition, analogous to the subject's difficulty in ignoring words while naming colors on the color-word task. In dissonance terms, it could be said that the words represent dissonant cognitions. It would seem then that facility in ignoring or withstanding the disruption provoked by the words in the color-word task extends to the area of beliefs. This strategy of dissonance reduction is less available to constricted smokers, and they are forced to either tolerate the dissonance or reduce their smoking.

It should be noted that there are certain difficulties in attempts to account for individual differences in dissonance experiments in terms of cognitive controls or other subject characteristics. These problems have been discussed at length by Brehm and Cohen (1962) and Glass (1965). Briefly, the basic point is that since dissonance can only be measured indirectly (i.e., via efforts to reduce it), it becomes difficult to specify whether the obtained individual differences are due to differences in the perception, tolerance, and/or manner and degree of dissonance reduction. This problem might eventually be solved by including subject variables which have specific relevance to different phases of the dissonance arousal-reduction sequence. On the basis of the present data, one cannot definitively indicate which aspect of the dissonant situation is related to the constricted-flexible-cognitive-control dimension. Tentatively, it is assumed that all smokers cannot very readily avoid exposure to (i.e., perception of) information linking smoking and cancer (Feather, 1963). Furthermore, there is no reason to assume that constricted subjects, who as a group smoke more than flexible subjects, are more tolerant of dissonance or experience less of it. Rather, it appears that the cognitive resolution via denial is simply

less available to constricted subjects so that constricted and flexible subjects differ in the manner of dissonance reduction.

It has been suggested (Gardner et al., 1959) that cognitive controls and defenses might involve similar structures which serve adaptive or defensive aims, respectively. Applying this conception to the present data, it is proposed that while resistiveness to intrusive cognitions manifests itself adaptively on the color-word task (i.e., is conducive to effective performance), it serves as a defense (perhaps denial or isolation of affect) on the smoking questionnaire insofar as it indicates a failure or unwillingness to acknowledge a reality.⁶ Constricted subjects, being more vulnerable to intrusive information, either have to tolerate the dissonance or resort to other ways of resolving the situation (e.g., decreased smoking). It is interesting to note, however, that in contrast to most other studies (e.g., Holt, 1960; Loomis & Moskowitz, 1958) this study showed that a generally more adaptive control (viz., flexible control) can, in certain situations, possibly have maladaptive effects (i.e., denial of the reality of a smoking-cancer link).

Relevance of Individual Differences in the Study of Dissonance

The present study is an example of how the assessment of stable cognitive characteristics can increase the predictive accuracy of the dissonance model. If the nonsmokers had been compared with the smokers without regard to the constricted-flexible distinction, there would have been no difference in mean belief. Granted an undersized sample of nonsmokers, there was not even a noticeable trend in the direction of the dissonance-theory prediction that smokers would exhibit greater denial. It might be argued that the increasing evidence linking smoking and cancer makes it difficult to deny such a link without flagrant disregard for reality. That this is not the case is indicated by the significant belief difference between constricted and flexible smokers.

Simply varying treatment conditions, then, could have readily obscured individual differences in the impact of a given condition and would have led to the erroneous conclusion that denial was no longer a feasible means of reducing dissonance based on smoking. This point is applicable to a recent study by Pervin and Yatko (1965), where negligible differences were found between smokers and nonsmokers on such variables as contact with and selective recall of relevant information, and estimates of the dangers of smoking. Assessment of relevant

⁶ Inquiry concerning the bases of subjects' belief might have been instructive in this regard.

subject variables might have produced clear differences among the smokers.

The selection of subject characteristics for assessment should be clearly related to the experimental situation. For example, the constricted-flexible-cognitive-control dimension is probably most relevant when the reality of the dissonant situation is unambiguous, for that is when individual differences in denial are likely to come into play. Thus, it probably would have been easier for constricted smokers to deny the unpleasant aspects of smoking than it was for them to deny the link between smoking and cancer. Similarly, at the time of Festinger's study, more than 10 years ago, the evidence linking cigarette smoking and lung cancer was less definitive, and the constricted-flexible distinction might not have been related to difference in belief at that time.

REFERENCES

- BAER, D. J. Smoking attitude, behavior, and beliefs of college males. *Journal of Social Psychology*, 1966, 68, 65-78.
- BREHM, J. W., & COHEN, A. R. *Explorations in cognitive dissonance*. New York: Wiley, 1962.
- EHRlich, D., GUTTMAN, I., SCHÖNBACH, P., & MILLS, J. Postdecision exposure to relevant information. *Journal of Abnormal and Social Psychology*, 1957, 54, 98-102.
- FEATHER, N. T. Cognitive dissonance, sensitivity, and evaluation. *Journal of Abnormal and Social Psychology*, 1963, 66, 157-163.
- FESTINGER, L. *A theory of cognitive dissonance*. Evanston, Ill.: Row, Peterson, 1957.
- GARDNER, R. W., HOLZMAN, P. S., KLEIN, G. S., LINTON, H. B., & SPENCE, D. P. Cognitive control: A study of individual consistencies in cognitive behavior. *Psychological Issues*, 1959, 1(4), 1-186.
- GLASS, D. C. Theories of consistency and the study of personality. In E. F. Borgatta & W. W. Lambert (Eds.), *Handbook of personality theory and research*. Chicago: Rand McNally, in press.
- HEIDER, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.
- HOLT, R. R. Cognitive controls and primary processes. *Journal of Psychological Researches*, 1960, 4, 1-8.
- HORN, D. Behavioral aspects of cigarette smoking. *Journal of Chronic Diseases*, 1963, 16, 383-395.
- KLEIN, G. S. Need and regulation. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1954*. Lincoln: University of Nebraska Press, 1954. Pp. 224-280.
- KLEIN, G. S., SCHLESINGER, H. J., & MEISTER, D. E. The Effects of personal values on perception: An experimental critique. *Psychological Review*, 1951, 58, 96-112.
- LOOMIS, H., & MOSKOWITZ, S. Cognitive style and stimulus ambiguity. *Journal of Personality*, 1958, 26, 349-364.
- OSGOOD, C. E., & TANNENBAUM, P. H. The principle of congruity in the prediction of attitude change. *Psychological Review*, 1955, 62, 42-55.
- PERVIN, L. A., & YATKO, R. J. Cigarette smoking and alternative methods of reducing dissonance. *Journal of Personality and Social Psychology*, 1965, 2, 30-36.
- WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received May 12, 1966)

Journal of Personality and Social Psychology
1967, Vol. 5, No. 4, 490-496

ANTICIPATION OF EXPOSURE TO PERSUASIVE MESSAGES AND BELIEF CHANGE

DEMETRIOS PAPAGEORGIS¹

University of British Columbia

150 undergraduates were told that they would listen to speeches advocating positions opposite to their initial beliefs and that the effect of these speeches on their opinions would be measured. Warnings about forthcoming speeches were given on 4 issues, and the effectiveness and reputability of their sources were varied orthogonally in order to test predictions from self-esteem theory. The topic and direction of argument of a 5th speech was announced, but Ss were told that they would not hear it. The inclusion of this condition allowed comparison between predictions derived from self-esteem and cognitive balance theories. Postwarning beliefs of Ss favor the balance interpretation of forewarning and suggest that belief change following warning may not be anticipatory. A compromise interpretation in terms of variable orientations toward belief expression is presented.

Most of the experimental research in opinion change is carried out with the purpose of the

¹ William J. McGuire made several helpful suggestions on an earlier version of this report.

The author was formerly at the University of Texas.

experiment disguised from the subjects. Implicit in this method is the assumption that the persuasive message will lose some or all its impact if the subject is told or discovers (i.e., is intentionally or unintentionally forewarned) that an attempt to influence his opinions is about

to be made. Experiments that have used forewarning in an enforced-exposure context have given only partial support to this assumption, and have also contributed other findings of considerable interest (Allyn & Festinger, 1961; Freedman & Sears, 1965; Kiesler & Kiesler, 1964; McGuire & Millman, 1965; McGuire & Papageorgis, 1962; Mills & Aronson, 1965).

Among these other findings are belief changes, in a direction opposite to the subject's initial position, that are found following the forewarning, but prior to any attack with the persuasive messages (McGuire, 1966; McGuire & Millman, 1965; McGuire & Papageorgis, 1962; Sears, Freedman, & O'Connor, 1964). The label "anticipatory" lowering or change of belief is applied often to this phenomenon (McGuire, 1966): the change is assumed to be a defensive response by the subject who, expecting to be gullible to the persuasive message and thus lose self-esteem, shifts his position closer to the position that he anticipates in the forthcoming attacking message. Depending on the extent of the shift and the position that the subject will eventually assign to the message, he may attain the postmessage luxury of perceiving himself as less gullible than he would have otherwise (Deutsch, Krauss, & Rosenau, 1962; McGuire & Millman, 1965).

In their recent study, McGuire and Millman obtained evidence that supports the anticipatory, self-esteem explanation of postwarning belief change. They found that forewarning about emotional unverifiable issues significantly changed preattack beliefs when these beliefs were compared to those expressed by subjects who had not been forewarned or who had been forewarned about technical, matter-of-fact issues; in addition, forewarning about technical issues did not change preattack beliefs. These findings support the self-esteem interpretation because opinion change in response to a factual communication is less threatening to one's self-esteem. The same investigators also hypothesized that forewarning of attack by a disreputable source would produce greater change in preattack beliefs than would forewarning of attack attributed to a reputable source; subsequent persuasion by the reputable source would be less injurious to self-esteem. This prediction was not supported. McGuire and Millman, however, pointed out that the source manipulation they used created differential reputability only within a context of high source effectiveness, and thus may have failed to maximize the difference in overall source persuasiveness needed for an adequate test of self-esteem predictions. The present study includes an attempt to overcome this difficulty. As suggested by McGuire and Millman, it introduces an orthogonal variation of source effective-

ness and reputability whereby conditions of high and low source effectiveness are paired with conditions of high and low source reputability. If these source manipulations are successfully induced, self-esteem theory would predict that preattack beliefs will show greater change following forewarning about a forthcoming message attributed to a source of (a) high effectiveness and (b) low reputability. Persuasion by *a* is more likely, while persuasion by *b* would be more threatening to self-esteem.

Postwarning belief change may also be derived from cognitive balance approaches (Heider, 1958; Newcomb, 1953). The forewarning informs the subject that somebody disagrees with him, thus causing a state of imbalance among cognitions and pressure to reduce this state. Belief change would then be a means of achieving better cognitive balance. This alternative derivation does not require that the subject expect actual exposure to the persuasive message; thus, a clearer test between the cognitive balance and self-esteem positions can be made if the experiment includes an additional condition in which the subject is informed of a belief contrary to his own, but is also assured that he will not be exposed to a persuasive message on that belief (cf. McGuire & Millman, 1965). The balance approach would predict belief change whether a subject expects exposure to a persuasive attack or not. Self-esteem theory, on the other hand, would predict belief change only if exposure is expected. In addition, if the balance interpretation is correct, the postwarning belief changes need not be anticipatory. The present experiment introduces a condition in which the subject is informed of a belief contrary to his own, but is also assured that no message on that belief will follow; this condition, hereafter called the "consistency control" or CC condition, in conjunction with the conditions where the subject expects a persuasive message to follow the forewarning, allows a test between the two derivations.

METHOD

The experiment was designed to assess the effect of forewarning on preattack beliefs under these six conditions: anticipated attack with a message attributed to a source of high effectiveness and high reputability (HH), anticipated attack by a high effectiveness and low reputability source (HL), anticipated attack by a low effectiveness and high reputability source (LH), anticipated attack by a low effectiveness and low reputability source (LL), no anticipated attack on an explicitly mentioned belief that is contrary to the subject's initial belief (CC), and no forewarning, a control condition (C).

Each subject participated in all six conditions, each condition measured by his belief on a different issue. Each issue was used with every experimental condition using different subjects. The se-

quence of presentation of the five issue-condition combinations that were mentioned to the subjects (one of the issues always serving as an unmentioned control) was controlled by presenting each of the combinations in every position in the sequence. The choice of obtaining data from each subject for all experimental conditions was based on several considerations, including the desire to maximize differential perception between the four source conditions. In addition, this choice made the CC manipulation both plausible in itself and comparable to the anticipated-message conditions. This design also made it economically possible to use six different issues, thus increasing the generality of the findings. The latter was also enhanced by the use of seven experimenters who, though all were males, varied in age and experimental sophistication and had formed no theoretical preferences.

Since the experiment dealt only with postwarning belief change, and not with the effect of warning on communication impact, no actual persuasive messages were presented to the subjects, though each subject was led to believe that four messages were forthcoming and was given the topic and direction of argument of a fifth message that supposedly would not be presented. Within the present experimental context, however, the forewarning itself can be considered as the persuasive communication. In these terms, the design of the experiment conforms to the "after-only" type (Hovland, Lumsdaine, & Sheffield, 1949).

The results of the experiment were evaluated by means of analysis of variance and two-tailed tests of statistical significance. Probability level for rejection of the hypothesis of no difference between means was set at .01.

Procedure

The subjects participated in small groups of from three to eight persons. Each of these groups was randomly assigned to one of the seven experimenters and to a particular sequence of one of the six issue-condition combinations. Since each issue was used in each of the six conditions (HH, HL, LH, LL, CC, and C), and since order of presentation was controlled by using five sequences of presentation (with HH, HL, LH, LL, and CC each appearing in all five positions in the sequence), 30 of these small groups were needed. The alternative of assigning individual subjects, rather than groups, to conditions would have made it impossible for the experimenter to read the materials aloud to the subjects and could also have created suspicion if the subjects discovered that the sequence and content of their materials differed.

Aside from these design variations, the experimental procedure was quite simple. After the subjects had assembled in the seminar-type room used for the experiment, the experimenter announced that he was measuring opinion change that resulted from listening to actual recorded speeches. A tape recorder, threaded with (blank) tape, was displayed prominently, and the experimenter stated that the

machine would be used to play the recorded speeches. The instructions also indicated that information about the topics of the speeches and the speakers would be given before any of the speeches were played; in addition, the subjects were told that they were to give their opinions about the topics before and after they had heard the speeches so that the experimenter could assess any changes in their opinions. Every effort was made to convince the subjects that the speeches would be forthcoming and that the objective of the experiment was as stated.

Following the initial instructions, the subjects were given descriptions of the topics and of the direction of argument of five speeches. In addition, speaker descriptions, in terms of effectiveness and reputability, were given for four of these speeches; the instructions repeatedly stressed not only that these four speeches would be played, but also that the fifth speech, for which no speaker description was given, had not been received and, consequently, would not be played. The descriptions of the five speeches were circulated in written form, and the subjects retained them for reference during the experiment; the experimenter also read them aloud in order to make sure that the subjects paid attention.

The experimenter announced next that, before listening to the speeches, the subjects were to rate the speakers and state their opinions on the issues. The speaker ratings were used as a measure of the effectiveness of the source manipulation. The opinion measure was the dependent variable.

The experiment was complete after the administration of the opinion measure. At this point, the subjects could have been told the purpose of the experiment and could have been informed that there was never any intent to play speeches to them. Since, however, data collection was not complete, the explanation of the experimental procedures had to be postponed. Instead, the apparently distraught experimenter invoked a variety of excuses for postponing the playing of the speeches; for example, the room had been scheduled for another purpose, the tape recorder would not function, and so forth.

Subjects

Undergraduates enrolled in first psychology courses served as subjects. Participation in experiments was required, although the choice of particular experiments was open to the students. The sample contained roughly equal numbers of males and females and was likewise representative of both lower- and upper-division students. A total of 157 subjects participated, but 7 subjects were eliminated at random from those issue conditions that totaled over 25 subjects. Thus, the final number of subjects used in the experiment was 150, with 25 subjects in each of the six issue-condition variations. The subgroups within these variations that received different sequences of presentation varied in size from a low of three to a high of eight subjects.

Materials

Issues. The issues were selected about 2 months prior to the experiment from the responses to a belief questionnaire by a sample of 191 persons similar to those subsequently used in the experiment. The questionnaire was similar to the belief questionnaire used as the dependent variable measure (see below). The items in this questionnaire included those used in the McGuire and Millman experiment in addition to other items that were expected to satisfy the criteria for issue selection. These criteria were extremeness and homogeneity of expressed belief. All items dealt with issues that appeared to be emotional and unverifiable, rather than with technical, matter-of-fact ones, since a greater postwarning impact had been demonstrated with the former variety of issue (McGuire & Millman, 1965).

The six issues called for by the experimental design were selected on the basis of the results of the early belief survey. Eighty-five percent or more of the sample expressed beliefs that were uniformly distributed within one or the other end of the true-false belief continuum for all six issues. These were:

1. A third world war is almost certain to occur in the very near future.
2. There will soon be a serious economic depression in the United States.
3. In the near future, the United States Government will find it necessary to reduce the freedoms now constitutionally guaranteed to citizens.
4. In the near future, Americans will have very little to say in their choice of a physician.
5. Soon, college graduates will have to undertake advanced graduate study before they are considered eligible for many positions open to them today.
6. The United States will contribute very little that is new to the exploration of outer space.

With the exception of 5, initial beliefs were distributed within the false end of the continuum. To facilitate analysis and presentation of results, 5 was transcribed in its mirror-image form; thus, for all six issues, low scores represent the subject's initial position, and movement toward the position advocated by the speeches is indicated by higher scores.

Forewarning. The forewarning was both general and specific to the issues. The general forewarning was effected in several ways, including the recruitment of subjects for an experiment labeled "Opinion Change," the instructions that stressed repeatedly that the objective of the experiment was to measure opinion change that would result from listening to persuasive speeches, the administration of the belief measure after the subjects were told that it was a "before" measure of their opinions, and the presence of the threaded tape recorder.

In addition to the general warning, specific warnings were given for five of the six experimental issues, the sixth issue serving as an unmentioned control. Each specific warning stated the topic of a talk and the direction of its argument. The direction of the argument was always opposite to the assumed initial stands of the majority of the subjects. Each subject was led to expect that he would

hear four of these talks and one given without description. Source names for the HH, HL, LH, and LL source conditions appeared in the first source. The fifth was assigned to the CC condition in which the subject was specifically and negatively informed that he would not hear the talk.

Source descriptions. Descriptions designed to make comparisons of high and low source effectiveness were prepared independently with knowledge of high and low source reputability. All the names were printed with one of the four effectiveness reputability conditions, though the high and low reputability conditions were not always matched.

High speaker effectiveness was manipulated by describing the speakers as learned and powerful leaders, noted writers, and generally as elite people who were presenting prepared speeches. Low speaker effectiveness was created by describing the speakers as inexperienced individuals speaking under less than ideal conditions, usually extemporaneously. Spoken reputation in the other hand was manipulated through the speakers' reputation as effective with one group, thus a federal judge, an ex-senator, a reformer, and a member of the President's Council of Economic Advisors were used as high speaker reputability conditions, while an extreme right wing politician and an official opponent for the Trade Union, and an official of the Teamsters Union represented examples of the people used for the low reputability condition. Within an experiment the two conditions involving high reputability (HH and LH) included, to the extent that this did not interfere with the internal consistency of the source description, the same information. Thus, high speakers were given the same fiction to make the same affiliation, and the same general character. The same was true for the two low reputability conditions (HL and LL), the two high effectiveness conditions (HH and HL), and the two low effectiveness conditions (LH and LL). Space limitations do not permit a complete reproduction of the 24 source descriptions.³ For purposes of illustration, the four descriptions used with Issue 1 (likelihood of a third world war) follow:

HH The Honorable James W. Bakstrap, deputy undersecretary of defense, Washington, D. C. Mr. Bakstrap is a noted spokesman for the Department of Defense; his talk has been delivered before a variety of audiences and has also been broadcast on the air with a good deal of success.

HL Joseph C. Brown, secretary of "Americans for Freedom," an organization identified with the extreme right wing of American politics. Mr.

³ A complete set of the source descriptions has been deposited with the American Documentation Institute. Order Document No. 9213 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D.C. 20540. Remit in advance \$1.25 for microfilm or \$1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

- Brown is a frequent and effective spokesman for Americans for Freedom. He has spoken to many groups, and his talk was broadcast over several radio stations.

LH: The Honorable James W. Bickstrup, deputy undersecretary of defense, Washington, D. C. Mr. Bickstrup is a highly regarded planner for the Department of Defense, but is not known for his public-speaking activities. His talk is actually a series of unrehearsed, off-the-cuff comments made to a radio reporter and, consequently, may lack in organization and style of presentation.

LL: Joseph C. Brown, secretary of "Americans for Freedom," an organization identified with the extreme right wing of American politics. Mr. Brown is not a frequent speaker; his talk is actually a series of unrehearsed, off-the-cuff comments made to a radio-station reporter and, consequently, may lack in organization and style of presentation.

The effectiveness and reputability of the sources assigned to the other five issues were manipulated in a similar fashion.

Following the presentation of the issues and their sources, the subjects rated each of the four speakers that they expected to hear in terms of the kind of speaker they expected the person to be, and the kind of person they thought each of the speakers was. These ratings were made on semantic differential scales. The scales "effective-ineffective" and "persuasive-unpersuasive" were used to measure the first type of rating. The second type of rating was made on the scales "disreputable-reputable" and "trustworthy-untrustworthy." The purpose of the ratings was to check on the success of the orthogonal source manipulation.

Belief measure. The dependent-variable belief measure consisted of a series of statements, including the six experimentally relevant items, each followed by a 15-interval graphic scale. The scale was labeled from left to right in intervals of three as "Definitely False," "Probably False," "Uncertain," "Probably True," and "Definitely True." Essentially this is the same scale that was used in several earlier studies (e.g., Papageorgis, 1963).

RESULTS

Source Manipulation

In four of the experimental conditions the subjects were led to anticipate communications attributed to sources that varied in effectiveness and reputability. It is necessary to show that the source descriptions did indeed create different perceptions of effectiveness and reputability. The subjects' ratings of the sources were used as indexes of their perceptions. Two semantic differential scales were used to measure perceived source effectiveness, and two similar scales were used as a measure of perceived reputability. Scores from these two scales were added, thus yielding

TABLE 1
MEANS AND STANDARD DEVIATIONS OF
POSTWARING BELIEFS

Condition	Issues					
	1	2	3	4	5	6
CC	6.76 (2.86)	5.68 (3.00)	6.12 (3.58)	6.32 (3.77)	7.04 (3.47)	3.24 (3.51)
HH	7.16 (2.99)	6.48 (2.94)	6.96 (3.74)	5.88 (3.49)	6.52 (2.98)	2.96 (2.07)
LH	5.96 (2.20)	6.60 (3.46)	6.24 (3.93)	6.72 (4.22)	6.96 (3.15)	3.56 (2.84)
HL	7.96 (2.88)	5.36 (2.94)	7.20 (4.16)	6.24 (4.14)	6.32 (3.08)	2.32 (1.65)
LL	6.40 (2.29)	5.52 (2.63)	6.72 (4.34)	5.04 (2.54)	5.92 (3.31)	2.20 (1.47)
C	6.32 (3.09)	6.04 (3.26)	5.16 (3.54)	5.48 (3.10)	4.04 (1.57)	2.04 (1.27)

Note.— $N = 25$ per cell. Standard deviations appear in parentheses.

an overall measure of each source dimension with a range from 2 to 14. The higher scores represent greater perceived effectiveness or reputability. The success of the source manipulation was evaluated by comparing mean effectiveness and reputability ratings for the high and low effectiveness and reputability conditions.

The mean comparisons showed that the source manipulations were successful. The mean for the high source reputability conditions was 11.97, while the mean for the low reputability condition was 8.44 ($p < .01$). Similarly, the mean for high speaker effectiveness was 11.22, and the mean for low effectiveness was 9.56, a difference also significant at the .01 level.

Belief Changes

The control mean belief for all six issues was 4.85. The means for the other conditions were: 5.99 (HH), 5.90 (HL), 6.01 (LH), 5.30 (LL), and 5.86 (CC). Except for the LL mean, all other experimental means differ significantly from the control mean ($p < .01$).

The main effect for issues was not significant. The Issue \times Condition interaction was significant ($p < .01$). The interaction is primarily the result of a significant difference between LL and C on the issue dealing with constitutional freedoms, and no significant differences between HL, LL, CC, and C on the issue of a forthcoming economic depression. Table 1 presents the means and standard deviations for all conditions and issues.

DISCUSSION

The principal conclusion suggested by the significant postwarning belief changes is that anticipation of a forthcoming persuasive message is not necessary for this kind of belief change. For five of the six experimental issues, the CC condition resulted in belief levels significantly above the control level. Moreover, the CC-induced belief levels are of the same magnitude as those induced by the anticipated message conditions that were attributed to sources with high perceived effectiveness or high perceived reputability. The results then favor a cognitive balance interpretation and suggest that postwarning belief changes may be viewed as adjustments in a person's beliefs that result from awareness that another person holds different beliefs. The present study generally replicates the McGuire and Millman (1965) findings in that (a) the warning causes preattack belief change, and (b) the source effects are negligible. At the same time the anticipatory interpretation of postwarning change given by these authors is not supported.

The finding of belief changes without anticipation of a forthcoming message and solely on the basis of information about the existence of a persuasive speech requires some comment. Two earlier attempts to induce belief change by communicating majority opinions were unsuccessful (Brock & Blackwood, 1962; Robbins, 1961). This discrepancy is most likely the result of differences in experimental procedure: the announcement of the existence of a persuasive speech may carry greater inconsistency-arousing potential than the knowledge of majority opinion of peers. In addition, neither of the previous studies was presented to the subjects as an opinion-change experiment. It is also possible to raise questions regarding the exact nature of the CC manipulation used in the present experiment. One objection may be that the subjects became confused as to which speeches they were scheduled to hear and which speech would not be presented. This possibility is unlikely since the instructions repeatedly indicated the specific issue-condition pairings, and the same materials were available to the subjects for reference throughout the experiment. A second possibility may be that the CC condition, because it was always presented with the anticipated speech conditions, may have acquired a status of psychological equivalence with the latter. In other words, the subjects were affected by the CC condition only because it was always paired in the design with anticipated speeches. The fact

that the LL condition was not effective may argue somewhat against this interpretation, but the possibility that the CC manipulation is effective only under conditions that create a set of expectancy for other persuasive communications remains open, subject to further experimentation, and suggests caution about possibly premature generalizations of the present finding.

Although self-esteem explanations of postwarning change in belief receive little support from the present results, evidence from other studies, such as the finding of change with emotional but not with factual issues (McGuire & Millman, 1965), appears quite compatible with the self-esteem position and is not easily accounted for in terms of balance theories. A somewhat different hypothesis that would reconcile existing findings and incorporate features of both balance and self-esteem theories may be needed. Such a hypothesis may be stated as follows: A person, when asked to state his belief about an issue, may do so after assuming either of two sets or orientations toward belief expression. With the first of these sets, the person responds as an individual and expresses whatever position he assumes he has at the time; under the second set, the person will respond only after he has weighed the significance of the item and the nature of his own position relative to a larger social context. With this second orientation, a person may, if his preresponse deliberations suggest it, express a different belief than he would under the first, individual set; he may express a belief that will be more in line with what he assumes or expects other people believe, or with his assessment of the extent of his factual knowledge about the issue, or with his feelings about how justified his belief is, and so forth.

Forewarnings, as well as many other kinds of advance information about an issue regardless of whether a person anticipates an attempt at persuasion, create tendencies that result in subsequent belief expression that includes the deliberative activity characteristic of the second, social context type of orientation. The social orientation may or may not result in modification of the individual set beliefs; if it does, the person will probably state a belief that is congruent with the advance information he has acquired, although it is conceivable that he may resist the information. The latter case amounts to an "immunization" or a "boomerang" effect of the forewarning, depending on whether forewarning is viewed as a defense-stimulating procedure or as a persuasive communication.

The term belief expression, rather than change or adoption, was used in the above analysis to suggest the need for caution in describing the effects of advance information on beliefs. The small belief change that follows warning may be no more than a fluctuation within the existing latitude of the person's belief (Sherif & Hovland, 1961). Everyday observation suggests that, for many beliefs, people retain the privilege of varying their stand in different contexts and circumstances while still remaining on the same side of the issue.

Forewarning, of course, does not exhaust the means of altering belief-expression sets. The present hypothesis suggests effects that may result from other orientation-directed manipulations. For example, will subjects change their stand if: (a) They are told that their beliefs will be quoted in a publication; or (b) if they are led to expect that their beliefs will affect the future course of an official policy or the fate of a television show; or (c) if they are simply informed that their beliefs will be measured more than once? Findings from relatively simple studies of this type would allow a more proper evaluation of the present hypothesis. They may also point out the need for greater caution in the instruction procedures used in many belief- and attitude-change studies.

REFERENCES

- ALLYN, J., & FESTINGER, L. The effectiveness of unanticipated persuasive communications. *Journal of Abnormal and Social Psychology*, 1961, 62, 35-40.
- BROCK, T. C., & BLACKWOOD, J. E. Dissonance reduction, social comparison, and modification of opinion. *Journal of Abnormal and Social Psychology*, 1962, 65, 319-324.
- DEUTSCH, M., KRAUSS, R. M., & ROSENBAUM, N. Dissonance or defensiveness? *Journal of Personality*, 1962, 30, 16-28.
- FREEDMAN, J. L., & SEARS, D. O. Warning, distraction, and resistance to influence. *Journal of Personality and Social Psychology*, 1965, 1, 262-266.
- HEIDER, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.
- HOVLAND, C. I., LUTMSDAINE, A. A., & SHEFFIELD, F. D. *Experiments on mass communication*. Princeton: Princeton University Press, 1949.
- KIESLER, C. A., & KIESLER, S. B. Role of forewarning in persuasive communications. *Journal of Abnormal and Social Psychology*, 1964, 68, 547-549.
- MCGUIRE, W. J. Attitudes and opinions. In, *Annual review of psychology*. Vol. 17. Palo Alto, Calif.: Annual Reviews, 1966. Pp. 475-514.
- MCGUIRE, W. J., & MILLMAN, S. Anticipatory belief lowering following forewarning of a persuasive attack. *Journal of Personality and Social Psychology*, 1965, 2, 471-479.
- MCGUIRE, W. J., & PAPAGEORGIS, D. Effectiveness of forewarning in developing resistance to persuasion. *Public Opinion Quarterly*, 1962, 26, 24-34.
- MILLS, J., & ARONSON, E. Opinion change as a function of the communicator's attractiveness and desire to influence. *Journal of Personality and Social Psychology*, 1965, 1, 173-177.
- NEWCOMB, T. M. An approach to the study of communicative acts. *Psychological Review*, 1953, 60, 393-404.
- PAPAGEORGIS, D. Bartlett effect and the persistence of induced opinion change. *Journal of Abnormal and Social Psychology*, 1963, 67, 61-67.
- ROBBINS, P. R. Immediate and delayed effects of social influence upon individual opinion. *Journal of Social Psychology*, 1961, 53, 159-167.
- SEARS, D. O., FREEDMAN, J. L., & O'CONNOR, E. F. The effects of anticipated debate and commitment on the polarization of audience opinion. *Public Opinion Quarterly*, 1964, 28, 617-627.
- SHERIF, M., & HOVLAND, C. I. *Social judgment*. New Haven: Yale University Press, 1961.

(Received May 5, 1966)

